BULLETIN

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

EARLE R. HEDRICK

DAVID R. CURTISS WILLIAM R. LONGLEY

WITH THE ASSISTANCE OF

HAROLD T. DAVIS HARRY W. KUHN

THEOPHIL H. HILDEBRANDT DERRICK N. LEHMER

JOHN R. KLINE HENRY L. RIETZ

CAROLINE E. SEELY

VOLUME XXXIX, NUMBER 8

AUGUST, 1933

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1933

PUBLISHED MONTHLY

Whole No. 407 $9.00 a Year

Entered as second class matter April 24, 1926 at the post office at Menasha, Wis., under the act of August 24, 1912.

AMERICAN MATHEMATICAL SOCIETY

Mail Address: 501 West 116th Street, New York, N.Y., for the Secretary, Treasurer, Librarian, and Clerk.

Office: Room 413, Physics Building, Columbia University, New York City.

President: Professor A. B. Coble, University of Illinois, Urbana, Ill.

Secretary: Dean R. G. D. Richardson, Brown University, Providence, R.I.

Associate Secretaries: Professor M. H. Ingraham, University of Wisconsin, Madison, Wis.; Professor J. R. Kline, University of Pennsylvania, Philadelphia, Pa.; Dean T. M. Putnam, University of California, Berkeley, Calif.

Treasurer: Professor G. W. Mullins, Barnard College in Columbia University, New York City.

Librarian: Professor R. C. Archibald, Brown University, Providence, R.I.

Clerk: Dr. Caroline E. Seely, 413 Physics Building, Columbia University, New York City.

Editor-in-Chief of the Bulletin: Professor E. R. Hedrick, University of California at Los Angeles, Los Angeles, Calif.

Managing Editor of the Transactions: Professor R. D. Carmichael, University of Illinois, Urbana, Ill.

Chairman, Colloquium Editorial Committee: Professor R. L. Moore, University of Texas, Austin, Tex.

Managing Editor of the American Journal of Mathematics: Professor Abraham Cohen, Johns Hopkins University, Baltimore, Md.

Representatives of the Society in the Division of Physical Sciences of the National Research Council: Professors A. B. Coble, E. R. Hedrick, and Dean R. G. D. Richardson.

Representatives on the Council of the American Association for the Advancement of Science: Professors Louis Ingold and M. H. Ingraham.

Chairman, Committee on Arrangements for the Semi-Centennial: Professor T. S. Fiske, Columbia University, New York City.

ON THE INFINITE SEQUENCES ARISING IN THE THEORIES OF HARMONIC ANALYSIS, OF INTERPOLATION, AND OF MECHANICAL QUADRATURES*

BY LEOPOLD FEJÉR

1. *Introduction.* The three mathematical theories indicated in my title are so extensive that I should naturally be unwilling within the bounds of a single discussion to give an outline of the totality of the relevant investigations. On the contrary I shall in each case bind myself to a portion of the corresponding theory. The investigations which I have in mind, and which I hope to be able to present to you, have been conducted almost entirely in the twentieth century. Even in this portion of the theory, however, so many brilliant contemporary mathematicians have collaborated that a considerable complex of investigations has resulted. Thus I shall select from this narrower field only a few results,—such, however, as are characteristic and have served as points of departure for further researches.

I shall therefore undertake to give only an outline of these dominant characteristic results, and shall accomplish this by exhibiting as clearly as possible the *single fundamental idea* which unites them. If I can succeed in the course of my lecture in making the investigations of the whole complex seem to you less diversified, I shall have achieved my goal.

2. *Fourier Series.* I begin my exposition with Fourier series. If $f(t)$ denotes an integrable real function of the real variable t, having the period 2π, then the constants

$$(1)\qquad \begin{cases} a_0 = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} f(t)\,dt, \\ a_n = \dfrac{1}{\pi}\displaystyle\int_0^{2\pi} f(t)\cos nt\,dt, \\ b_n = \dfrac{1}{\pi}\displaystyle\int_0^{2\pi} f(t)\sin nt\,dt, \end{cases} \qquad (n = 1, 2, 3, \cdots),$$

* An address delivered by invitation of the authorities of the Century of Progress Exposition and the American Association for the Advancement of Science at the World's Fair in Chicago, June 21, 1933, before the American Mathematical Society and Section A of the A.A.A.S.—The author is indebted to Professor W. A. Hurwitz for the preparation of the English version of this address.

are known as the Fourier constants of $f(t)$, and the infinite series

$$(2)\qquad a_0 + a_1 \cos x + b_1 \sin x + \cdots + a_n \cos nx + b_n \sin nx + \cdots = \sum_{n=0}^{\infty}(a_n \cos nx + b_n \sin nx)$$

is called the Fourier series of the function f. The sum of the first $n+1$ terms of this series is clearly

$$(3)\qquad s_n(x) = \frac{1}{2\pi}\int_0^{2\pi} f(t)\{1 + 2\cos t \cos x + 2 \sin t \sin x + \cdots + 2 \cos nt \cos nx + 2 \sin nt \sin nx\}\,dt.$$

In most cases, however, it is convenient to note that

$$1 + 2 \cos t \cos x + 2 \sin t \sin x + \cdots + 2 \cos nt \cos nx + 2 \sin nt \sin nx = \frac{\sin (2n+1)\frac{t-x}{2}}{\sin \frac{t-x}{2}},$$

so that we have for the so-called partial sum of the Fourier series of index n for the value x the Dirichlet formula

$$(4)\qquad s_n(x) = \frac{1}{2\pi}\int_0^{2\pi} f(t)\frac{\sin (2n+1)\frac{t-x}{2}}{\sin \frac{t-x}{2}}\,dt.$$

Now let us form the arithmetic mean $S_n(x)$ of the partial sums $s_0(x), s_1(x), \cdots, s_n(x)$, so that

$$(5)\qquad S_n(x) = \frac{s_0(x) + s_1(x) + \cdots + s_n(x)}{n+1}.$$

From the obvious formula

$$(6)\qquad S_n(x) = \frac{1}{n+1}\,\frac{1}{2\pi}\int_0^{2\pi} f(t)\{(n+1) + n\cdot 2\cos t \cos x + n\cdot 2 \sin t \sin x + \cdots + 1\cdot 2 \cos nt \cos nx + 1\cdot 2 \sin nt \sin nx\}\,dt,$$

by use of the identity

$$(7)\quad (n+1) + n\cdot 2\cos t\cos x + n\cdot 2\sin t\sin x + \cdots + 1\cdot 2\cos nt\cos nx + 1\cdot 2\sin nt\sin nx = \left[\frac{\sin (n+1)\frac{t-x}{2}}{\sin\frac{t-x}{2}}\right]^2,$$

follows

$$(8)\quad S_n(x) = \frac{1}{2\pi(n+1)}\int_0^{2\pi} f(t)\left[\frac{\sin (n+1)\frac{t-x}{2}}{\sin\frac{t-x}{2}}\right]^2 dt, \qquad (n = 0, 1, 2, \cdots).$$

If to each value x of an interval is made to correspond a numerical value of f, f is called a function of x. If to each function f of a set of functions a numerical value A is made to correspond, then A is said to be obtained from f by a functional operation (briefly *operation*). If we now think of x as fixed, and correspondingly write s_n, S_n for $s_n(x)$, $S_n(x)$, then obviously s_n and S_n are each formed by an operation on $f(t)$:

$$(9)\qquad s_n = s_n[f],$$

$$(10)\qquad S_n = S_n[f].$$

These operations are both ordinary linear functional operations. An essential difference between them is, however, that the operation $S_n[f]$ is *positive*, while the classical operation $s_n[f]$ does *not* possess this property. *An operation $A[f]$ is called positive, provided $A[f] \geqq 0$ whenever $f(t) \geqq 0$ throughout the interval of values of t in which the function $f(t)$ is considered.*

I published this result in 1900. It is remarkable that from the classical *indefinite* operation $s_n[f]$, the *definite* (positive) operation $S_n[f]$ can be obtained by a very slight modification without any limit process. The sharpness of the contrast between these two types of linear operation, as viewed at the time of my investigation, undoubtedly increased the long existing interest of mathematicians in this distinction.

So much has been published concerning the meaning of the operation $S_n[f]$ of the arithmetic mean, and various kinds of more delicate considerations, further generalizations, and applications, that I must unfortunately omit all mention of them, in accordance with the plan of this lecture. I must thus neglect a mass of material, to which I have myself often endeavored to contribute, and which has been enriched by a long list of profound articles by other mathematicians.

3. *Further Remarks on Fourier Series.* I continue, still remaining within the domain of Fourier series. As we have essentially seen already, the positiveness of the operation $S_n[f]$ follows immediately from the simple fact that the ordinary arithmetic means of the partial sums of the series

$$(11) \qquad 1 + 2\cos\theta + 2\cos 2\theta + \cdots + 2\cos n\theta + \cdots$$

are all non-negative for every real value of θ. Very recently I have made the observation that for the series

$$(12) \quad 0 + \sin\theta + 2\sin 2\theta + 3\sin 3\theta + \cdots + n\sin n\theta + \cdots$$

the arithmetic means of third order $S_n^{(3)}(\theta)$ of the partial sums $S_n^{(0)}(\theta) = s_n(\theta)$, are all positive in the interval $0 < \theta < \pi$ (except of course $S_0^{(3)}(\theta) \equiv 0$). For the means of orders 0, 1, and 2 of the series (12) this assertion is *not* valid. I have further noted,—and this seems to lie somewhat deeper,—that for the series

$$(13) \qquad \sin\theta + 3\sin 3\theta + 5\sin 5\theta + \cdots$$

even the arithmetic means of second order of the partial sums are all non-negative in the interval $0 < \theta < \pi$. (In this case even more detailed statements can be made.) For the means of orders 0 and 1 of the series (13) this assertion is *not* valid.

To these properties of the series (12) and (13) again correspond in a certain sense *positive linear functional operations* on the pure Fourier sine-series or the pure cosine-series of a function $f(x)$. But I shall not exhibit these operations here, nor discuss the numerous interesting consequences of their positiveness.

As, however, this is an entirely new direction of investigation, I shall allow myself to mention just one characteristic result.

Let the function $f(x)$ be positive and convex (or even not con-

cave) upward everywhere in the interval $0<x<\pi$. In order to have a particularly interesting special case, I assume further that the curve $y=f(x)$ has the property of symmetry,

$$f(\pi-x) = f(x) \quad \text{for} \quad 0 \leqq x < \frac{\pi}{2}.$$

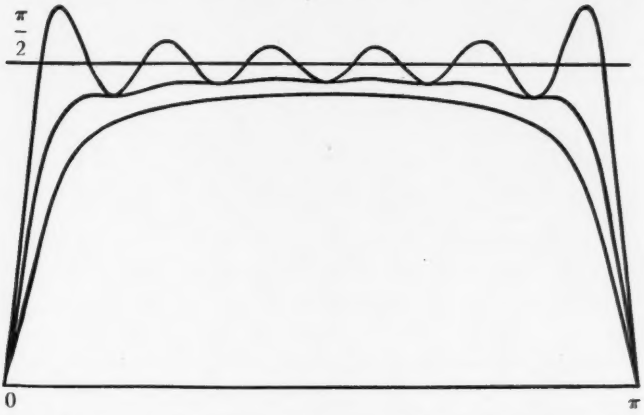
The Fourier sine-series of the function $f(x)$ has the form

$$(14)\quad f(x) \sim b_1 \sin x + b_3 \sin 3x + \cdots + b_{2\nu-1} \sin (2\nu-1)x + \cdots,$$

where

$$(15)\qquad b_{2\nu-1} = \frac{2}{\pi}\int_0^{\pi} f(t) \sin (2\nu-1)t\,dt, \quad (\nu = 1, 2, 3, \cdots).$$

As an immediate consequence of the statement about the series (13) we have the following theorem: all arithmetic means of second order of the partial sums of the Fourier sine-series (14) for the function $f(x)$ are positive and convex upward in the interval $0<x<\pi$. The arithmetic means of orders 0 and 1 on the other hand do *not* in general possess this property of convexity.



The accompanying figure* illustrates this theorem in the simplest imaginable special case, in which $f(x)$ is constant in the

* I am indebted to Mr. I. Raisz for the careful drawing of this figure.

interval $0 \leqq x \leqq \pi$ (thus $y=f(x)$ is non-concave upward). The Fourier sine-series of the function

$$f(x) = \frac{\pi}{2}, \qquad (0 < x < \pi), \tag{16}$$

is

$$f(x) = \frac{\pi}{2} = 0 + \frac{\sin x}{1} + 0\frac{\sin 2x}{2} + \frac{\sin 3x}{3} + 0\frac{\sin 4x}{4} + \frac{\sin 5x}{5} + \cdots . \tag{17}$$

The upper curve of this figure represents the partial sum of index 12 of this sine-series (17), that is,

$$S_{12}^{(0)}(x) = \sum_{\nu=1}^{6} \frac{\sin (2\nu - 1)x}{2\nu - 1} .$$

The middle curve represents the arithmetic mean of first order $S_{12}^{(1)}(x)$ and the lower curve the Cesàro mean of second order $S_{12}^{(2)}(x)$ of the series (17). The interesting curves $S_{12}^{(0)}(x)$, $S_{12}^{(1)}(x)$ are already found in the book of H. S. Carslaw on Fourier Series in connection with the discussion of the famous Gibbs phenomenon, which arises in the case of the partial sums $s_n(x) = S_n^{(0)}(x)$ but disappears for the means of first order $S_n^{(1)}(x)$. If we now examine the three curves from the point of view of convexity-concavity, we observe that both $S_{12}^{(0)}(x)$ and $S_{12}^{(1)}(x)$ are composed of arcs, which are alternately convex and concave upward. On the other hand the new curve $S_{12}^{(2)}(x)$, the lower curve, which represents the Cesàro mean of second order of index 12 for the series (17), is convex upward in the whole interval $0 < x < \pi$.

Theorems such as the preceding illustrate from a new angle the "smoothing-out effect" on the Fourier series of the process of forming repeated means. I should like to conclude this consideration with only the further observation that the smoothing-out effect of repeated formation of means is very notable in another connection. If

$$w = f(z) = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n + \cdots \tag{18}$$

is a power series which is regular and assumes each of its values

only once in the circle $|z|<1$, this property is not in general shared by the partial sums

$$(19)\quad S_n^{(0)}(z)=s_n(z)=c_0+c_1z+\cdots+c_nz^n,\ (n=0,1,2,3,\cdots),$$

for the whole unit circle. But the arithmetic means of second or third order $S_n^{(2)}(z)$, $S_n^{(3)}(z)$ of the partial sums (19) maintain the property of the original function (i.e. take on each of their values only once in the whole unit-circle), provided the power series $c_0+c_1z+c_2z^2+\cdots$ belongs to a certain special, but remarkable, subclass of the totality of series having the specified property.

4. *Laplace Series.* If $S_n^{(0)}[f]=s_n[f]$ denotes the partial sum of index n of the Laplace series for a function $f(\theta, \phi)$, integrable over the unit sphere, then it is well known that for a fixed point (θ, ϕ) of the unit sphere $S_n^{(0)}[f]$ represents an indefinite linear operation on f. The operation $S_n^{(1)}[f]$, which is defined by the arithmetic means of first order of the Laplace series, is also indefinite. But the operation $S_n^{(2)}[f]$ is *positive.*

I obtained this result in 1908. It is now possible to give a proof as follows. Obviously, we have only to show that the means of second order for the series

$$(20)\quad P_0(\cos\theta)+3P_1(\cos\theta)+5P_2(\cos\theta)+\cdots+(2n+1)P_n(\cos\theta)+\cdots$$

are positive for $0<\theta<\pi$. This is accomplished at one stroke by the use of Mehler's formula for the Legendre polynomial $P_n(\cos\theta)$:

$$(21)\quad P_n(\cos\theta)=\frac{2}{\pi}\int_\theta^\pi \frac{\sin(2n+1)\frac{t}{2}}{[2(\cos\theta-\cos t)]^{1/2}}dt,$$

since, as previously stated, the means of second order of the series

$$(22)\quad \sin\frac{t}{2}+3\sin 3\frac{t}{2}+5\sin 5\frac{t}{2}+\cdots+(2n+1)\sin(2n+1)\frac{t}{2}+\cdots$$

are positive for $0<t<\pi$.

I shall not dwell further on positive operations in the domain of Fourier and Laplace series. But in any case I must at least refer to the two important positive operations introduced by D. Jackson. Both are closely related to our first example of a linear operation, that of the arithmetic mean, in which occurs the "kernel" $(\sin nt/\sin t)^2$. Jackson's first operation is used for the approximation (by trigonometric polynomials) to functions subjected to certain restrictions. His second operation is used for the trigonometric approximation to arbitrary continuous functions; it consists in a method of interpolation, and is closely related to the operations which we shall consider next.

5. *Interpolation.* In the present part of this lecture I shall concern myself with interpolation, and in particular, I shall indicate briefly how the theory of interpolation can utilize the idea of the positive linear functional operation.

If $x_1, x_2, \cdots, x_n$ denote arbitrary distinct real numbers, then the Lagrange interpolation formula is

$$L(x) = y_1 l_1(x) + y_2 l_2(x) + \cdots + y_n l_n(x). \tag{23}$$

This formula represents the polynomial of degree at most $n-1$, which assumes the arbitrarily given values $y_1, y_2, \cdots, y_n$ respectively for the values $x_1, x_2, \cdots, x_n$. Here

$$l_k(x) = \frac{\omega(x)}{\omega'(x_k)(x - x_k)}, \quad (k = 1, 2, \cdots, n), \tag{24}$$

where

$$\omega(x) = C(x - x_1)(x - x_2) \cdots (x - x_n), \qquad (C \neq 0). \tag{25}$$

Usually $y_1, y_2, \cdots, y_n$ are the ordinates of a given curve $y=f(x)$ for the abscissas $x=x_1, x_2, \cdots, x_n$; that is, $y_1=f(x_1)$, $y_2=f(x_2), \cdots, y_n=f(x_n)$. In this sense, therefore, the classical Lagrange formula defines for a fixed value of x a linear functional operation $L[f]$. For *no* choice of the abscissas $x_1, x_2, \cdots, x_n$ (if $n \geqq 2$), however, is this operation *positive* (for any fixed value of x different from $x_1, x_2, \cdots, x_n$). This is because each of the so-called fundamental polynomials $l_1(x), l_2(x), \cdots, l_n(x)$ of the Lagrange interpolation formula changes its sign $n-1$ times.

Let us now examine from this point of view the simplest so-called Hermite interpolation formula

$$X(x) = \sum_{k=1}^{n} y_k h_k(x) + \sum_{k=1}^{n} y_k' \mathfrak{h}_k(x). \tag{26}$$

This represents the polynomial of degree at most $2n-1$ which for the values $x_1, x_2, \cdots, x_n$ assumes respectively the values $y_1, y_2, \cdots, y_n$ and whose derivative correspondingly assumes the values $y_1', y_2', \cdots, y_n'$. Here $h_1(x), h_2(x), \cdots, h_n(x)$ are the so-called "fundamental functions of the first kind" for the Hermite interpolation formula (26):

$$h_k(x) = \left(1 - \frac{\omega''(x_k)}{\omega'(x_k)}(x - x_k)\right)(l_k(x))^2, \quad (k = 1, 2, \cdots, n), \tag{27}$$

and $\mathfrak{h}_1(x), \mathfrak{h}_2(x), \cdots, \mathfrak{h}_n(x)$ are the "fundamental functions of the second kind":

$$\mathfrak{h}_k(x) = (x - x_k)(l_k(x))^2, \quad (k = 1, 2, \cdots, n), \tag{28}$$

where $\omega(x)$ and $l_k(x)$ have the same meaning as before.

If we now merely glance at the Hermitian fundamental functions $h_k(x)$ and $\mathfrak{h}_k(x)$, we see at once that they have, in contrast to the Lagrangian fundamental functions $l_k(x)$, a kind of tendency to be positive. I shall make this statement more precise.

The fundamental function $\mathfrak{h}_k(x)$ is equal to the square of the polynomial of $(n-1)$th degree $l_k(x)$ multiplied by the linear function

$$w_k(x) = x - x_k. \tag{29}$$

Hence $\mathfrak{h}_k(x)$ changes its sign exactly once, at the interpolation point x_k.

The fundamental function $h_k(x)$ is equal to the square of $l_k(x)$ multiplied by the linear function

$$v_k(x) = 1 - \frac{\omega''(x_k)}{\omega'(x_k)}(x - x_k). \tag{30}$$

Thus $h_k(x)$ also can change its sign only once, at the vanishing point X_k of the linear function $v_k(x)$,

$$X_k = x_k + \frac{\omega'(x_k)}{\omega''(x_k)}. \tag{31}$$

The point X_k is however, under all circumstances, different from x_k, since $v_k(x_k)=1$. If the interpolation points

$$x_1, x_2, \cdots, x_n \tag{32}$$

are given, the points

$$X_1, X_2, \cdots, X_n \tag{33}$$

are uniquely determined. I shall call X_k the *conjugate point* corresponding to the interpolation point x_k.

Henceforward let the interpolation points $x_1, x_2, \cdots, x_n$ lie in the interval $-1 \leqq x \leqq 1$, and let x range over this interval only. We then obtain the following result:

The Hermite fundamental polynomials of the first kind are non-negative in the whole interval of interpolation $-1 \leqq x \leqq 1$, *if and only if all the conjugate points* $X_1, X_2, \cdots, X_n$ *lie outside the interval of interpolation* $-1 < x < 1$.

If we now consider the linear functional operation

$$H(x) = \sum_{k=1}^{n} f(x_k) h_k(x), \tag{34}$$

whose meaning is easily discerned, we can make the following restatement:

A necessary and sufficient condition for the positiveness of the linear operation (34) *for every fixed value of* x *in the interval* $-1 \leqq x \leqq 1$ *is that the interval* $-1 < x < 1$ *be free from conjugate points.*

By the assumption of the positiveness of a linear functional operation (the operation (34)) we are thus led to the classification of all point-systems $x_1, x_2, \cdots, x_n$ of the interval $-1 \leqq x \leqq 1$ into two classes. To the first class belong those systems $x_1, x_2, \cdots, x_n$ for which the conjugate points $X_1, X_2, \cdots, X_n$ all lie outside the interval $-1 < x < 1$; to the second class belong all other systems.

The importance of this distinction which rests on our postulate of positiveness of the operation lies primarily in the fact that those point-systems $x_1, x_2, \cdots, x_n$ which are useful for

interpolation and are of greatest mathematical interest belong to our first class.

For instance let $x_1, x_2, \cdots, x_n$ be the Tschebyscheff abscissas, which present themselves from so many modes of approach. They may be obtained by describing a semicircle having as diameter the interval from -1 to $+1$ on the x-axis, dividing this semicircle into n equal arcs, and projecting the points of bisection of these arcs on the x-axis.

In this special case it is easily seen that the point X_k conjugate to x_k is $X_k = 1/x_k$; i.e., X_k is the harmonic conjugate to the interpolation point x_k with respect to the points $-1, +1$. Since the harmonic conjugates to $x_1, x_2, \cdots, x_n$ all lie outside the interval $(-1, +1)$, it is clear that the Tschebyscheff abscissas $x_1, x_2, \cdots, x_n$ belong to our first class of point-systems.

As a second example we choose the point-system which was introduced by Gauss in his famous monograph on mechanical quadrature by parabolic interpolation. The points of interpolation $x_1, x_2, \cdots, x_n$ are now the roots of the equation

$$P_n(x) = 0, \tag{35}$$

where $P_n(x)$ denotes the Legendre polynomial of index n. It is readily seen that in this case $X_k = (x_k + 1/x_k)/2$; that is, the conjugate point X_k is half-way between the point of interpolation x_k and its harmonic conjugate. But the point of bisection of the interval between two conjugate harmonic points always lies outside the interval between the fundamental points. The set of Legendre-Gauss abscissas thus belongs to our first class of point-systems.

More generally the roots $x_1, x_2, \cdots, x_n$ of the equation

$$J_n(\alpha, \beta, x) = 0 \tag{36}$$

always belong to the first class, provided that $J_n(\alpha, \beta, x)$ denotes the so-called nth Jacobi polynomial and the parameters α, β satisfy the inequalities $0 \leqq \alpha \leqq \frac{1}{2}$, $0 \leqq \beta \leqq \frac{1}{2}$. For $\alpha = \beta = \frac{1}{2}$ this set of roots is that of Gauss, for $\alpha = \beta = \frac{1}{4}$ that of Tschebyscheff. I call attention to a third special case of exceptional interest: $\alpha = \beta = 0$. In this case the points of interpolation are $-1, +1$, and the $n-2$ roots of the equation $P'_{n-1}(x) = 0$ where $P_{n-1}(x)$ denotes the Legendre polynomial of index $n-1$.

For infinite sequences of Lagrange and Hermite interpolation polynomials of a function, provided the point-systems always belong to the first class, it is now possible to state very general theorems, all of which can be proved with surprising ease. But I cannot dwell further on this important subject. I proceed to my third and last topic—mechanical quadrature by parabolic interpolation.

6. *Mechanical Quadrature.* Let $f(x)$ be a function defined and Riemann-integrable in the interval $-1 \leqq x \leqq 1$. If $x_1, x_2, \cdots, x_n$ again denote n distinct points of the interval $-1 \leqq x \leqq 1$, then

$$L(x) = \sum_{k=1}^{n} f(x_k) l_k(x) \tag{37}$$

is the corresponding Lagrange parabola and the value of the integral

$$Q = \int_{-1}^{+1} L(x)dx = \sum_{k=1}^{n} f(x_k) \int_{-1}^{+1} l_k(x)dx \tag{38}$$

is called the corresponding *mechanical quadrature*. The factors

$$\lambda_1 = \int_{-1}^{+1} l_1(x)dx, \lambda_2 = \int_{-1}^{+1} l_2(x)dx, \cdots, \lambda_n = \int_{-1}^{+1} l_n(x)dx, \tag{39}$$

the integrals of the Lagrange fundamental functions over the interval of quadrature, which depend only on the abscissas $x_1, x_2, \cdots, x_n$, I call the Cotes numbers. It is clear that the quadrature formula

$$Q = Q[f] = \sum_{k=1}^{n} \lambda_k f(x_k) \tag{40}$$

again represents a linear functional operation; and I remark at once that it is in general indefinite, i.e., there are both positive and negative numbers among the Cotes numbers $\lambda_1, \lambda_2, \cdots, \lambda_n$. This is true, for instance, if for an infinite set of values of n the points $x_1, x_2, \cdots, x_n$ divide the interval $(-1, +1)$ into n equal parts. In this connection I may cite the important investigations of Uspensky and Pólya, and I should like to remark that the

results of the latter have given a new impulse to this direction of study.

From what has preceded one is led almost involuntarily to a new classification of point-systems $x_1, x_2, \cdots, x_n$ of the interval $(-1, +1)$. In the first class are placed those systems $x_1, x_2, \cdots, x_n$ for which the Cotes numbers $\lambda_1, \lambda_2, \cdots, \lambda_n$ are all non-negative, in the second class all other systems.

Does there exist even one single point-system $x_1, x_2, \cdots, x_n$ which belongs to the first class? This question is answered in the affirmative by the classical result of Gauss, Christoffel, and Stieltjes; for the Legendre abscissas $x_1, x_2, \cdots, x_n$ the Cotes numbers are all positive, i.e., the operation of quadrature is positive. I have recently discovered that the Cotes numbers are positive also for the Tschebyscheff abscissas; for the Jacobi abscissas if $\alpha=\beta=0$ or $\alpha=\beta=\frac{3}{4}$ or $\alpha=\frac{1}{2}, \beta=0$, and for other systems as well. Subsequently Szegö in an original fashion has treated the question of positiveness of the Cotes numbers for the most general Jacobi systems.

Numerous theorems of very general character can now be proved with the utmost ease, dealing with point-systems for which the operation of quadrature is definite, and with the corresponding quadratures themselves. I shall mention one such theorem: if we have an infinite sequence of such point-systems, consisting of more and more points, then the corresponding quadratures converge to $\int_{-1}^{+1} f(x)dx$ as $n=\infty$, if $f(x)$ is bounded and Riemann-integrable in the interval. This theorem is a far-reaching generalization of the theorem of Stieltjes, which enunciates the same result on convergence for the special case of the Legendre-Gauss point-system. But I consider it remarkable that it has apparently been overlooked hitherto that such special systems, as for instance that of Tschebyscheff, also belong to our first category of point systems.

We have divided point systems $x_1, x_2, \cdots, x_n$ of the interval $-1 \leqq x \leqq 1$ sharply into two classes in connection with interpolation; we have again divided these systems sharply into two classes in a different way in connection with mechanical quadrature. May these two classifications be characterised in any other way? What relation subsists between the two classifications? I content myself with a mere mention of these problems.

Just one word more about the case of quadrature. For

Fourier and Laplace series and for interpolation the original classical expressions (for instance the partial sums of the series, the Lagrange polynomials of the interpolation) are indefinite; only among the modified expressions (arithmetic means, Hermite interpolation polynomials, etc.) do we find those which are positive. For quadrature, on the other hand, we have seen that the original quadrature, obtained by means of the classical Lagrange parabola, is *positive*, provided the abscissas of the quadrature belong to a certain well defined class.

7. *Conclusion.* I have come to the end of my lecture, to which I might as well have given the following title: On the significance of the idea of the positive linear operation for harmonic analysis, interpolation, and quadrature. I should like to add just one brief remark, relating to the whole subject-matter of the lecture. In each case we were concerned with a set of linear operations, from which we selected those that were positive, in order to arrive at certain goals appropriate to the various theories. Is it actually necessary, in order to arrive at these goals, to demand that the operations be positive? By no means. In order to obtain necessary and sufficient conditions, a consideration of absolute values is demanded, as in the familiar work of Lebesgue in breaking the path for the study of singular integrals. Then why have we concerned ourselves so particularly with positive operations? Because the most elementary operations which we meet at the outset in the three fields are fortunately positive. And for another reason as well: if we do not seek complete generality, but subject the operations to certain natural supplementary restrictions, then positiveness of the operations is both necessary and sufficient.

THE UNIVERSITY OF BUDAPEST

SOME MATHEMATICAL ASPECTS OF THE NEW MECHANICS*

BY TULLIO LEVI-CIVITA

1. *The Antecedent of Any Mechanical Conception. The Moving Point.* The mental pictures of natural phenomena are of very different kinds. Among the simplest and most basic of these mental pictures we may include the kinematical model of the motion of a geometrical point. Such a picture is an abstraction, i.e., an ideal construction, which may be obtained conveniently through axiomatic treatment of geometry and time.

The use of this model for real bodies necessitates, of course, a further *physical* postulate, or axiom, quite independent of previous kinematical assumptions. But it is so intuitively evident, in common-life scale, that no thinker has till now felt the need of elucidating or criticizing its adoption. We may designate it as the "axiom of substantiality of matter," or more briefly, of Descartes, if we agree to give this precise meaning to the famous assertion of Descartes "tout se fait par figure et par mouvement." Indeed these words have been rather interpreted as generally alluding to a mechanical explanation of nature. Such an explanation, for a great body of facts (including terrestrial and celestial motion), was offered by Newton and his successors, and prevailed undisputed in the 18th century.

2. *Autonomous Phenomenological Theories.* In the following century new classes of phenomena were detected and thoroughly investigated, especially the propagation of heat and electromagnetism, for which autonomous theories were established which were entirely satisfactory as representations of experience and highly valuable from a mathematical point of view, but which were entirely or almost independent of mechanical framework. I need hardly mention the classical treatises of Fourier and Maxwell. Such phenomenological, even somewhat detached,

* An address delivered, by invitation of the authorities of the Century of Progress Exposition and the American Association for the Advancement of Science, in Chicago, June 20, 1933, before a joint meeting of the American Mathematical Society and Section A of the A.A.A.S.

theories seemed at the end of the last century (and may perhaps still seem) excellent as far as they offer a quantitatively exact description of observed circumstances, and are otherwise capable of being used for systematic predictions.

3. *Relativity and Efforts toward Unified Formulation of all Physics.* A little later Einstein's relativity showed, in a wonderful though unexpected way, an intimate linkage between laws of motion, gravitation, and geometrical optics. Weyl, Eddington, Einstein himself, and many, many others were led to search for a more general unified view which would include electromagnetism and therefore physical optics.

4. *Quantum Mechanics. Early Stage (Bohr-Sommerfeld).* Meanwhile atomic physics began, this name belonging to the imposing congeries of investigations—permit me to name only Rutherford and Millikan—on radiation, rays, and isolated particles which verified the existence of minute constituents of electricity and matter. Up to a certain point the usual principles were still applicable, permitting rough estimates, but the systematic use of Planck's quantum of action h demanded a modified attitude towards the classical laws of motion. The differential equations expressing them were not changed, but the constants of integration, or rather some of them (always including energy), were deprived of continuity, and supposed to assume only discrete values (integral multiples of the lowest). By regarding the atom of hydrogen as made up of a single negative electron revolving round a positive nucleus, Bohr succeeded in presenting to the astonished world a mechanical model of spectroscopic lines (Balmer's, Lyman's, and Paschen's series) of very high accuracy.

More generally the theory of adiabatic invariants (Ehrenfest, Epstein, Burgers), systematically developed by Sommerfeld, led to an exact or approximate model of other atoms (in concept we may say of all other atoms), with the same degree of accuracy.

This was hardly accounted for in the classical work of Sommerfeld, *Atombau und Spektrallinien*, when a more severe criticism of old facts and some new discoveries led to new foundations of quantum mechanics.

5. *The Failure of the Basic Postulate of Kinematical Substantiality.* We shall speak of the new mechanics in a few minutes;

but I wish now to call your attention to a singular event, the most curious, in my opinion, that has ever been recorded in the evolution of scientific points of view. As you know, it was only in the first years of the present century that the real existence of small particles like electrons and atoms was experimentally ascertained. Formerly the great majority of scientists were doubtful, or at least reserved in opinion, concerning the question "to be or not to be" of atoms, which was then metaphysical in the literal meaning of the word, i.e., beyond the means of control available at that time. But no one had suspected that such things, if they existed, did not obey the cartesian postulate of substantiality, nor that for the description of the main phenomena connected with them we would be compelled to give up the most ancient and familiar device for the study of the physical world: the merely kinematical concept of a moving geometrical point.

In fact new experiments have on the one hand afforded irrefutable evidence for the existence of something—atom, electron, or photon—having individual characteristics, at a certain place (representable by a geometrical point), at a given instant t. But on the other hand it was ascertained that these entities have so little consistency that they are essentially modified, sometimes even (in a certain sense) annihilated, by our observations. If such an observation has taken place at a given instant t, a further observation is influenced: the latter does not give a sufficient account of, say, the electron, inasmuch as the electron has been strongly affected by the former observation.

6. *Heisenberg's Principle of Indeterminacy.* Analogous difficulties occur, of course, if we try to make observations at two very near instants, or, as a limiting process, to measure at the same instant position and speed. Furthermore a careful discussion of the (mainly) optical devices furnishing these evaluations brought Heisenberg to the discovery of his concrete principle of indeterminacy (or uncertainty), which approximates the second law of thermodynamics in affirming some impossibility in our physical world: the spontaneous passage of heat to a hotter body, in thermodynamics, and, in atomic physics, the measurement of both position and velocity with absolute accuracy. Obviously the absolute accuracy alluded to has usually

only the potential meaning of an accuracy which might theoretically be increased at will. More precisely this indeterminacy principle states, for one degree of freedom, that if x designates the space coordinate of an electron (or photon), m its mass, v its velocity, and $p=mv$ its momentum, the inaccuracies Δx, Δp are unavoidably of such amounts as to render their product

$$\Delta x \cdot \Delta p > h, \tag{1}$$

where $h=(6.5)10^{-27}$ in C.G.S. units is the Planck's constant (elementary quantum of action).

For ordinary particles of matter in bulk, even very minute, say m as small as 10^{-3} gr. (a milligram-mass), the inequality (1), since $p=mv$, gives for the kinematical indetermination

$$\Delta x \cdot \Delta v > h/m > (6.5)10^{-24} \text{ cm.}^2/\text{sec.}$$

This is quite compatible with the greatest exactness available or desirable in macromechanics. But the matter is quite different for much smaller particles like electrons. The order of magnitude of m is then 10^{-27} gr., therefore h/m becomes a value of the order of 6.5, and the product of kinematical uncertainties $\Delta x \cdot \Delta v$ has to be expected to be considerably greater than unity:

$$\Delta x \cdot \Delta v > 1, \tag{1'}$$

which deprives the corresponding physical observations of any interest. In fact, the linear dimensions of an electron being of the order of 10^{-13}, a not illusory Δx must be less than this, and then (1′) requires

$$\Delta v > 10^{13} \text{ cm./sec.},$$

which is more than a hundred times the velocity of light, $3(10^{10})$. Because of such an indeterminacy no useful information may be obtained about velocity.

In conclusion, atoms, electrons, photons do exist, since several phenomena grant physical evidence of them, but the representation of them as moving particles (or even as a system of moving bodies) is not adequate, because the indications which real experimentation may supply along this line are too vague.

7. *Digression on Rays as Filiform Flow of Electricity.* If, however, the elementary kinematical (and in consequence any dy-

namical) description must be abandoned as far as a *single* electron is concerned, Heisenberg's principle does not entirely exclude another model capable of furnishing *statistical* information on real observable average effects.

For this purpose we must imagine not a single electron but a great many, apparently forming a current, e.g., a ray. To explore the field of velocity interior to the ray (which may be some centimeters in length), we may take, or think of taking, an experimental measure of velocity, say, in every cubic micron of the space occupied by the ray. Since Δx is here 10^{-4} cm., formula (1′) gives for the measured velocity an inaccuracy certainly greater than 10^4 cm./sec., but with a possibility of becoming less than, say, 10^5, i.e., 1 km./sec. This is a very satisfactory approximation for velocities comparable to that of light (i.e., for velocity over one tenth that of light). It follows from this estimate that the treatment of electric radiations (β-rays and the like) as a continuous flow is quite compatible with the principle of indeterminacy, and therefore even today may be used satisfactorily as a theoretical scheme to investigate such average properties of rays (e.g., form deviations, forces exerted on and by them) for which a continuous or discrete distribution of charges (electrons) are practically indistinguishable.

The logical scheme is the following: no physical reality may be attached (on an atomic scale) to the motion of a single electron; on the other hand a *field* theory of the continuous flow of such particles (essentially a local description in the most elementary Eulerian form) may be fairly well supplied by concrete experiences. Now, still relying on these experimental measures, there is nothing to prevent us from substituting in our mathematical representation for the Eulerian point of view that of Lagrange, that is, to resolve the flow into (hypothetic) elementary moving particles, or even to replace their continuous flow by a discrete bombardment of very minute projectiles, occurring frequently enough to produce (on the above approximation) the same mean effects. And we may even suppose that such projectiles have just the mass of a single electron. *In this way it becomes possible, and very interesting, to give a positive meaning to the motion of a great many electrons or photons forming rays of different kinds.*

In this connection I may perhaps be allowed to add a per-

sonal remark. As you certainly remember, the realization that β-rays were composed of electrons (charges of negative electricity without material support) was reached in 1902 by Kaufmann's experiments, which were in surprising agreement with a suggestive ballistic theory of Abraham (whose rigid electron was a mathematical and mechanical refinement of a simpler relativistic model of Lorentz). This conclusion, as far as the purely electrical constitution is concerned, was afterwards verified by other facts. But, as we have seen, by Heisenberg's principle, no ballistic scheme has, in this range, direct physical meaning. However, in 1909 I proposed an asymptotic theory of continuous flow which leads to the same physically controlled consequences (and to one more not yet tested) and which, unlike the former, is consistent even today.*

8. *Validity of Classical Physics for Estimates of Mean Values.* Some other facts concerning mean values are satisfactorily explained by this very elementary process of averaging through continuous flow. The Compton effect belongs to this class; and we may in this way understand the success of the very elementary *corpuscular* explanation (based on conservation of both momentum and energy) offered by Compton himself and Debye. Actually, as previously observed, the uncertainty principle prevents us from treating photons or electrons as *isolated* particles; but the diffusion of X-rays in a gas by electrons, which gives rise to the Compton effect, is just recognizable as a continuous flow. Then a further treatment of this flow as discontinuous and energetically quantized is justified as indicated in §7, and brings us again to the elementary explanation just mentioned. From this the very instructive assumption may be made that, *as far as observable means are concerned, principles and therefore formulas of classical mechanics are still available*, at least to within a certain degree of precision. For a further refinement, and a nearer approach to phenomena on an individual scale, other devices must be sought.

9. *The New Principles of Quantum Mechanics and Wave Mechanics Inferred from a Typical Case (Hydrogen Atom).* A direct analysis of experimental data, especially in spectroscopy, where

* *Teoria Asintotica delle Radiazioni Elettriche*, Rendiconti della Accademia dei Lincei, vol. 18 (1st sem. 1909), pp. 83–93.

only lines (frequencies) and intensities of these lines are observable, was claimed by Born, Jordan, and Heisenberg. They succeeded in building a satisfactory theory involving only an enumerable set of directly measurable parameters, and based, from a mathematical point of view, on the algebra of matrices.

A quite different theory, equivalent as regards positive results, but much closer, at least in the analytical scaffolding, to the old captivating models of mathematical physics, was soon after proposed and elaborated by Schrödinger.

He first took into consideration the hydrogen atom, searching for a partial differential equation of common type (hence linear, self-adjoint of the second order) admitting levels of energy (quantized values of energy) as characteristic values.

Schrödinger's guiding principle has been to express just in this form de Broglie's association of waves with ordinary corpuscular dynamics, the wave function ψ being strictly correlated with that principal function V of Hamilton, which would belong to the electron of a hydrogen atom, if corpuscular nonrelativistic treatment were allowed.

The result is remarkably simple. Let

$$H(p \mid x) = E \tag{2}$$

be the equation of the first order defining V, where p generally replaces the set

$$p_1 = \partial V/\partial x^1,\ p_2 = \partial V/\partial x^2,\ \cdots,\ p_n = \partial V/\partial x^n,$$

and x the set

$$x^1,\ x^2,\ \cdots,\ x^n$$

of the independent variables; in the present case n has to be put $=3$.

Schrödinger's operator S to be applied to ψ is simply $H-E$ itself, provided that in H every p_k has the operational meaning $(h/(2\pi i))\partial/\partial x^k$. The linear equation in V being thus

$$S\psi \equiv [H\{(h/(2\pi i))\partial/\partial x \mid x\} - E]\psi = 0, \tag{3}$$

the correct energy levels appeared as characteristic values of (3), and moreover, assuming

$$\psi = e^{2\pi i V/h}, \tag{4}$$

the Hamilton-Jacobi equation (3) follows for V, up to the first order of approximation with respect to h.

10. *First Attempt at a Mathematical Correlation between Waves and Particles. Born's Appeal to the Calculus of Probability.* The last conclusion was for a while regarded as capable of furnishing a well defined association between the motion of a particle in the field considered and a wave-packet surrounding the particle and perceptible only in the immediate neighborhood of the paticle itself. But a further study has shown that difficulties arise on presenting in detail any such correlation; above all it does not harmonize with Heisenberg's indeterminacy.

To suppress fundamental discrepancies of this sort, Born made a suggestion of wide philosophical application. Instead of connecting the wave-function ψ with some corpuscular entity, belonging to or associated with actually observed facts, he proposed an interpretation borrowed from statistical mechanics. More exactly, he substituted for Schrödinger's first assumption that $|\psi|^2$ should be proportional to a *real* electrical density, a less binding and highly fruitful probability interpretation, namely, that $|\psi|^2 dS$ is available as a distribution function, in the sense that $|\psi|^2 dS$(dS element of space) must be interpreted as the probability for the electron [or rather for an electron whose energetic and spectroscopic apparencies are correctly described by (2) or (3)] to be present in dS at the moment considered. This implies of course that $\int_S |\psi|^2 dS$ (the probability that the electron finds itself somewhere in the whole space) is a certainty, that is $=1$, a condition which is easily fulfilled by means of the constant factor which, on account of the linearity of (3), certainly remains at our disposal.

As in the kinetic theory, once in possession of the law of distribution, we are able to obtain probable values of any physical quantity, connected statistically with a very large number of others having the same macroscopic behavior.

11. *Relativistic Wave-Equation (for a Single Particle Moving in a Given Field).* According to Einstein's geodetic principle, the motion of a particle whose mass is m and electric charge e, under gravitational and electromagnetic forces of any kind, may be represented by a canonical system. Its Hamiltonian is*

* For the deduction of this Hamiltonian from the Einstein-De Donder variational postulate see Appendix I.

$$(5) \qquad H = (1/2)\sum_{0}^{3}{}_{\mu\nu} g^{\mu\nu}(p_\mu - e\phi_\mu/c)(p_\nu - e\phi_\nu/c) - m^2c^2/2,$$

where $g^{\mu\nu}$ denotes the (contravariant) metric tensor, ϕ_μ the covariant electromagnetic potential, referred, like e, to electrostatic units; c the light velocity in vacuo, and p_μ the momentum, connected with the space-time coordinate x^μ. As usual, we shall regard x^0 as the time variable, x^1, x^2, x^3 as space variables. In the absence of a gravitational field, the coordinates may be supposed pseudo-cartesian, and then the $g^{\mu\nu}$ have, like $g_{\mu\nu}$, their pseudo-euclidean values:

$$\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{matrix}$$

The canonical system, having (5) as Hamiltonian function, admits in any case the integral $H = \text{const.}$; we must take more strictly

$$(6) \qquad H(p \mid x) = 0,$$

as a consequence of the transformation which gives canonical form to the original principle of variation.

Schrödinger's relativistic equation may be written at once, extending the formal rule of §9: substitution of the operator $(h/(2\pi i))\partial/\partial x^\mu$ for p_μ. The only generalization consists in replacing the ordinary $\partial/\partial x^\mu$ by the covariant derivation ∇_μ (with reference to $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$). We have, therefore, putting for brevity

$$(7) \qquad h' = h/(2\pi),$$

omitting, as does Einstein, the symbol for summation, and multiplying the two sides by $-2h'^2$, Schrödinger's relativistic equation

$$(8) \qquad S\psi \equiv [g^{\mu\nu}(\nabla_\mu - ie\phi_\mu/(ch'))(\nabla_\nu - ie\phi_\nu/(ch')) + m^2c^2/h'^2]\psi = 0,$$

sometimes known as Gordon's equation.

In the absence of any (gravitational and electromagnetic) field, referring to pseudo-cartesian coordinates and putting

$\psi = e^{i(mc+E/c)x^0/h'}\Psi$, where Ψ is independent of x^0 and E is a constant energy, we have simply

$$-\sum_1^3{}_\mu \partial^2\Psi/(\partial x^\mu)^2 + \{m^2c^2 - (mc + E/c)^2\}\Psi/h'^2 = 0,$$

or even, if E is very small in comparison with mc^2,

$$\sum_1^3{}_\mu \partial^2\Psi/(\partial x^\mu)^2 + 2mE\Psi/h'^2 = 0,$$

the non-relativistic equation of Schrödinger.

12. *Spin. Pauli-Dirac Wave-Vector.* For a short time it was believed that (in the foreseen range) equation (8), which had been used with imposing success, should be definitive. But, while fine structure was satisfactorily included, striking observations were soon made concerning spin effect and magnetic disturbances (anomalous Zeeman-effect), which did not agree with the consequences of (8) as far as spectroscopic displacements are concerned.

A very interesting synthetic reasoning along the line of group theory brought Pauli to introduce the needed correction as Goudsmith and Ullenbeck had already done for Sommerfeld's corpuscular theory. Pauli substituted for (8) a system of two equations involving two wave-functions. A further improvement was made by Dirac through a thorough analysis and original formal translation of physical assumptions. This author has succeeded, we may say, in splitting up Pauli's differential equation of the second order into four equations of the first order with as many unknown functions $\psi_0, \psi_1; \psi_2, \psi_3$. This quadruple may be interpreted as a *vector* ψ in space-time. Dirac and his followers did not make this interpretation. They gave to the set of *wave-functions* $\psi_\mu (\mu = 0, 1, 2, 3)$ another interpretation deduced from the theory of representation of the Lorentz group and the so-called *spinors*.

In terms of Dirac's ψ's the real quantity

$$\sum_0^3{}_s \psi_s \psi_s^*$$

(star denoting passage to the conjugate imaginary) is to be interpreted as (probable) electric density in ordinary three-dimensional space.

Dirac's equations proved to be in agreement with the more refined experiences just mentioned, reproducing, if spin-correction is neglected, Schrödinger's equation as resolvent of the second order.

Dirac's line of thought, in both geometric and algebraic points of view, is strictly related to ordinary (pseudo-euclidean) space-time and to pseudo-cartesian coordinates, leaving, therefore, gravitation unaccounted for.

13. *General Coordinates. Gravitational Field. Criticism of Matrix Equations.* General methods of absolute differential calculus allow, of course, translation to general coordinates of any system given in particular variables. Many authors applied themselves to this task; a more recent attempt was made by Schrödinger,† De Donder and Miss Dupont,‡ and myself.§ Following Schrödinger I have worked out the most natural (in a certain sense even uniquely determined) extension not only to general coordinates, but even to any relativistic ds^2, including gravitational actions.

To reach this extension I have been forced (like Schrödinger and many others) to introduce a lattice, i.e., a system of four mutually perpendicular congruences of lines, in space-time. This lattice is not an intrinsic feature of the physical phenomenon; it has only an auxiliary role, and must therefore disappear from final equations, if properly treated. An illustration of a similar circumstance in well known more elementary instances is shown in Appendix II.

It appeared from my calculations that in the generalized Dirac's equations a residual influence of the auxiliary lattice is not to be avoided, and this is quite unreasonable, the lattice having been introduced as a mathematical tool without any reference to the atomic events.

Indeed Dirac's original equations for relativistic electrons in pseudo-euclidean space (i.e. in absence of gravitation) involve a set of 4-rowed numerical matrices. Their geometric counter-

† *Diracsches Elecktron im Schwerefeld*, Berichte der Preussischen Akademie, 1932, pp. 105–128.

‡ *Généralisation relativiste des équations de Dirac*, Académie Royale de Belgique, vol. 18 (1932), pp. 596–602.

§ *Diracsche und Schrödingersche Gleichungen*, Berichte der Preussischen Akademie, 1933.

part is a particular pseudo-cartesian lattice, but the application of the Lorentz transformation group to the whole system easily shows that by change of lattice the equations remain substantially unaltered. That is not the case in general relativity. Therefore, Dirac's equations must in my opinion be abandoned, without, it is understood, giving up the main progress realized by them in the domain of restricted relativity, which was to account for spin and magnetic perturbations.

14. *Vector Equations of Schrödinger's Type with Additional Term.* The equations I have recently proposed with the above aim can be briefly described as Schrödinger's vector equation (of the second order) with correcting term. It links together the 4 (covariant) components ψ_μ of a vector ψ, in the following way:

$$S\psi_\mu + \chi_\mu{}^\nu\psi_\nu = 0 \qquad (\mu = 0, 1, 2, 3), \tag{9}$$

S being Schrödinger's operation (8) and $\chi_\mu{}^\nu$ a field tensor (function of the x's) defined by

$$\chi_\mu{}^\nu = Cg_{\mu\rho}\epsilon^{\rho\nu\sigma\tau}F_{\sigma\tau}, \tag{10}$$

where C is a constant, $F_{\sigma\tau}$ the (antisymmetric) electromagnetic tensor, and $\epsilon^{\rho\nu\sigma\tau}$ the contravariant ϵ-tensor of Ricci, whose elements are 0 if the indices ρ, ν, σ, τ are not all different, and, on this supposition, are equal to $1/(-g)^{1/2}$, where $g = \|g_{\mu\nu}\|$, or to $-1/(-g)^{1/2}$ according as the permutation $\rho\nu\sigma\tau$ is of even or odd order with respect to the fundamental permutation 0123.

Equation (9) may be compared with the second-order consequences of Dirac's equations in pseudo-cartesian coordinates. Assuming

$$C = e/(2ch'), \tag{11}$$

we may verify that the two sets are identical for both typical cases of a magnetic or an electric field acting alone. For a general electromagnetic field, as well as for superposed effects of electromagnetism and gravitation, the two systems are no longer equivalent.†

Herein lies, therefore, the possibility of discriminating, by some experiment, between the two as to which is to be preferred.

† Professor De Donder has kindly communicated to me his belief that, using his results, one may reach complete equivalence.

Meanwhile it is to be observed that, under the general circumstances supposed here, the vector ψ (whose components are ordinarily complex) is related to probability as follows: if ψ_ν^* denotes the complex conjugate of ψ_ν, put

$$P = g^{\mu\nu}\psi_\mu\psi_\nu^*,$$

which is always a positive quantity, giving for a real vector ψ its square length. We assume that PdV_4 is the probability for the electron being in an elementary volume (more precisely hypervolume)

$$dV_4 = (-g)^{1/2}dx^0dx^1dx^2dx^3$$

of space-time manifold.

With the metric tensor $g_{\mu\nu}$, the elementary volume dS of a surface $x^0 = \text{const.}$ has the expression

$$(-gg^{00})^{1/2}dx^1dx^2dx^3.$$

Our probability may therefore be written

$$PdSdx^0/(g^{00})^{1/2},$$

or

$$PcdS/(g^{00})^{1/2}$$

for unit time (x^0 meaning ordinary time multiplied by c).

15. *Waves and Motion of Particles Mathematically Connected with a General System of Partial Differential Equations.* According to Hugoniot and Hadamard, any normal system of partial differential equations,† for example (9), in the four functions $\psi_0, \psi_1, \psi_2, \psi_3$, determines uniquely, by a very simple rule, a *first-order* partial differential equation for an unknown function $z(x^0, x^1, x^2, x^3)$ attributing a more coarse wave appearance to the refined phenomenon actually governed by the normal system (9). For our system (9) the equation defining z reduces to

$$H = g^{\mu\nu}p_\mu p_\nu/2 = 0, \qquad (p_\mu = \partial z/\partial x^\mu), \tag{12}$$

† A system is said to be normal if it involves as many equations as unknown functions, and is solvable with respect to certain highest derivatives. See my little volume (translated by G. Lampariello) *Caratteristiche dei Sistemi Differenziali e Propagazione Ondosa* (Bologna, Zanichelli, 1931), or the French translation (Paris, Alcan, 1932).

or, solved with respect to p_0,

$$p_0 = -H(p_1, p_2, p_3 \mid x^0, x^1, x^2, x^3), \tag{13}$$

about which, without writing H in full, it will be sufficient to say that H is necessarily homogeneous of the first degree with respect to p_1, p_2, p_3. Its solution z, or better the surfaces σ (in general variable with the time x^0) defined by

$$z(x^0, x^1, x^2, x^3) = \text{const.}, \tag{14}$$

are possible *wave fronts*, in the sense that any of the σ (and no other surfaces) may separate two regions in which different ψ's are solutions of (9), the two sets coinciding on the boundary.

If, as in the present case, the differential system verified by the ψ's is of the second order, then not only the functions themselves, but also their first derivatives, must be continuous across the σ, discontinuities affecting on the other hand second and higher derivatives.

Considerations which, at least in the ordinary pseudo-euclidean case, are well known, show that the equation $H=0$ has a very simple geometrical and physical meaning. It expresses the fact that a wave front, regarded as a moving surface in the surrounding three-dimensional space, advances (every element $d\sigma$ normally to itself) with the velocity of light. This displacement of any one of these elements $d\sigma$ gives rise to very expressive, though purely fictitious, point motions. It results from the general theory of characteristics that the point motions are regulated by the canonical system

$$dp_\mu/d\tau = -\partial H/\partial x^\mu,\ dx^\mu/d\tau = \partial H/\partial p_\mu, \qquad (\mu = 0, 1, 2, 3), \tag{15}$$

with any independent variable τ (and the numerical specification $H=0$); or in particular ($\tau = x^0$), by

$$dp_\mu/dx^0 = \partial H/\partial x^\mu,\ dx^\mu/dx^0 = -\partial H/\partial p_\mu, \qquad (\mu = 1, 2, 3), \tag{15'}$$

whose trajectories, as Cauchy's characteristics of (12), are the so called *bicharacteristics* of (9).

As you see, we have been able to associate in a very simple manner with (9)—as would be the same for more complicated physical processes—an *undulatory face* (waves of discontinuity) as well as a *corpuscular* one. This duality was first inferred by de Broglie with very striking intuition. It may be regarded,

thanks to the mathematical theory just quoted, as a very general law of natural philosophy.

16. *Specification of the Dualism, Waves-Particles, for Schrödinger's Equations. De Broglie's Wave-Length. Conclusion.* The extensive statement of the preceding section is, however, rather abstract and agnostic. It affirms, for every physical theory, the admission of waves of discontinuity, mathematically resolvable into motion of discrete particles. But it gives no criterion for concrete coupling of the two single appearances. Attempts in this direction have been made by de Broglie himself, Schrödinger (§10), Fermi and Persico, Maggi, and other scientists, leading to the conviction that no one-to-one correspondence between a wave (or a wave-packet) and a particle may be established without violating (as pointed out in §10) the principle of indeterminacy.

We may, however, enlarge upon the question, and (as in §7 for a rough treatment of electrons as electric current) attempt correspondences, not between individuals, but between sets of waves on the one hand and corpuscles on the other hand. To do this we must bear in mind some fundamental remarks concerning electrons or, analytically, the system (9):

(a) If it is not a single wave, then a wave train of discontinuity, passing across a little region ΔS in a little time Δt, as defined by (14), may be regarded as roughly equivalent to a set of electrons, every one of which may possibly be captured within ΔS and Δt.

(b) We shall admit, as in §7, that such a set of electrons is indistinguishable from a continuous flow, and that this flow is connected by (14) to the ordinary Hamiltonian scheme of analytical mechanics, i.e., equivalent to the motion of a great number of discrete particles according to the canonical system (15), or (15′).

This is not in contradiction with the analogous, but more elaborate, rule due to Schrödinger, in §11. We started then from the Hamiltonian (5) suited for ordinary particles of mass m and charge e. The actual Hamiltonian is $\mathfrak{H}$ of (15), or more symmetrically, in space-time manifold,

$$H \equiv g^{\mu\nu} p_\mu p_\nu / 2 \tag{12}$$

with the (integral) specification $H=0$ (see §15 and Appendix III). Now H is nothing but a limiting case of (5) for particles very minute with respect both to mass and charge. In fact, drop in (5) all terms containing a factor m or e and you have (12).

By such assumptions we have attributed a statistical significance to the general dualism of waves and particles, analytically expressed by the mere fact that every normal system (especially Schrödinger's) possesses both characteristic surfaces and bicharacteristic lines.

(c) To obtain in our case concrete physical results, as de Broglie did, we still need yet one more postulate; and that is the attribution to the hypothetical particles, correlated with waves and therefore controlled by (15), of the exact mass m belonging to each electron.†

This enables us (see Appendix III) to get the very remarkable formula

$$\lambda = h/((g^{00})^{1/2}mv), \tag{16}$$

where λ is the wave length of undulatory phenomena eventually accompanying electronic rays, m the mass of any of the electrons forming the ray, v the (mean) velocity of the flow, h Planck's constant, and g^{00}, as in §14, the pure time-like component of the contravariant metric tensor.

In restricted relativity, with reference to pseudo-cartesian coordinates, $g^{00}=1$, and (16) is precisely the relation discovered by de Broglie, before its experimental proof by Davisson and Germer, G. P. Thomson, and Rupp.

These hints may suffice, I hope, to justify my conviction that the analytical theory of characteristics of partial differential systems furnishes not only a sure and general frame work for the mathematical expression of the double, undulatory and corpuscular, appearance of any physical event, first foreseen and claimed by de Broglie; but even a statistical (if not as at first sought for, individual) connection between the two orders of things.

† This statement is not a consequence of (b), because according to the principle of indeterminacy, the continuous flow admitted there may not be regarded, from the physical standpoint, as made up of electrons: In fact by (b) we only introduced flow; the resolution of it into discrete particles, each of a given mass m, is a very natural, but independent hypothesis (postulate, in logical scheme).

APPENDIX I

Symmetrical Treatment of the Einstein-DeDonder Principle of Variation. The principle of variation governing the motion of a particle of mass m and charge e in a general (gravitational and electro-magnetic) field, is

$$\delta \int \{mcds + (ef_d/c)\} = 0, \tag{I.1}$$

where, in general coordinates,

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu, \tag{I.2}$$

and f_d means the Pfaffian

$$f_d = \phi_\mu dx^\mu, \tag{I.3}$$

whose coefficients are the covariant components of the electromagnetic potential.

The function to be varied in (I.1) may obviously be multiplied by an arbitrary constant, without modification of the content of variational condition. The form adopted is a result of the desire to assure the exact dimensions of an action. Independently from this specification,

$$mcds + (ef_d/c)$$

is homogeneous of the first degree with respect to the differentials dx; hence we may introduce an independent variable under the sign at will; e.g., the time-like x^0, or, more symmetrically, an auxiliary τ. Denoting by a dot derivatives with respect to τ, and putting

$$\mathfrak{x} = g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu/2, \qquad f = \phi_\mu\dot{x}^\mu, \tag{I.4}$$

we may present the variational equation (I.1) under the form

$$\delta \int \{mc(2\mathfrak{x})^{1/2} + (ef/c)\}d\tau = 0. \tag{I.5}$$

At any rate it is to be remembered† that, owing to the first-degree homogeneity, the four Lagrangian equations arising from

† See, for instance, Levi-Civita and Amaldi, *Lezioni di Meccanica Razionale*, vol. II, second part (Bologna, Zanichelli, 1927), p. 524.

(I.5) are not independent, and therefore not sufficient to determine the four x's as functions of τ; a circumstance which could be foreseen, as this auxiliary variable τ had been arbitrarily introduced. Hence we may add to (I.5) another relation, intended to fix the parameter τ.

Excluding for the present eventual extremals of zero-length, i.e. solutions of (I.1) along which ds^2 is identically $=0$, we shall find it convenient to take just $d\tau = ds$, or by (I.2) and (I.4),

$$2\mathfrak{x} = 1. \tag{I.6}$$

On account of this assumption, if we put for brevity

$$L = mc\mathfrak{x} + (ef/c), \tag{I.7}$$

(I.5) may be replaced by

$$\delta \int L d\tau = 0, \tag{I.8}$$

which, together with (I.6), is entirely equivalent to (I.5), (I.6).

The advantage of this transformation is that L is no longer homogeneous of the first degree in the $\dot{x}$'s. The Lagrangian system corresponding to (I.8) becomes of normal character in the $\ddot{x}$'s; and Hamilton's reduction to canonical form may be performed as usual, through the introduction of moments p_μ, conjugate to the $\dot{x}^\mu$, by the formulas

$$p_\mu = \partial L / \partial \dot{x}^\mu,$$

which, in virtue of (I.7), (I.4), takes the form

$$p_\mu = mcg_{\mu\nu}\dot{x}^\nu + \{e\phi_\mu/c\}, \qquad (\mu = 0, 1, 2, 3). \tag{I.9}$$

To get the Hamiltonian $H(x\,|\,p)$, we have only to put

$$H = p_\mu \dot{x}^\mu - L, \tag{I.10}$$

eliminating the $\dot{x}$'s by means of (I.9).

The resolution of (I.9) gives

$$\dot{x}^\mu = g^{\mu\rho}(p_\rho - e\phi_\rho/c)/(mc), \tag{I.11}$$

where every $g^{\mu\rho}$ is reciprocal to $g_{\mu\rho}$, i.e., the $g^{\mu\rho}$ are complementary minors (with due sign) of the $g_{\mu\rho}$ in the determinant of $\|g_{\mu\rho}\|$, divided by the determinant itself.

The expression (I.7) of L is broken up into the term $mc\mathfrak{x}$, homogeneous of the second degree in the $\dot{x}$'s, and the term ef/c, homogeneous of the first degree. Accordingly

$$p_\mu \dot{x}^\mu = (\partial L/\partial \dot{x}^\mu)\dot{x}^\mu$$

reduces, by Euler's theorem, to

$$2mc\mathfrak{x} + (ef/c),$$

and the Hamiltonian (I.10) to

$$H = mc\mathfrak{x},$$

where it remains only to replace the $\dot{x}$'s by (I.11). From the original expression (I.4) of $\mathfrak{x}$, we get finally

(I.12) $$\begin{aligned} H &= mc\mathfrak{x} \\ &= [(mcg_{\mu\nu}) \cdot g^{\mu\rho}\{p_\rho - (e\phi_\rho/c)\}/mc \cdot g^{\nu\sigma}\{p_\sigma - (e\phi_\sigma/c)\}/(mc)]/2 \\ &= g^{\mu\nu}\{p_\mu - (e\phi_\mu/c)\}\{p_\nu - (e\phi_\nu/c)\}/2(mc), \end{aligned}$$

which is the announced form.

The additional condition (I.6) gives

(I.13) $$H = mc/2,$$

which must be associated to the canonical system

(I.14) $$\dot{p}_\mu = -\partial H/\partial x^\mu, \quad \dot{x}^\mu = \partial H/\partial p_\mu, \qquad (\mu = 0, 1, 2, 3),$$

in order to get a differential system completely equivalent to (I.6), (I.8), or to the original variational principle (I.5)together with (I.6).

As H does not involve explicitly the independent variable τ, the differential system (I.14) admits the classical integral

$$H = \text{const.}$$

We see therefore that the canonical form just arrived at has the advantage, in addition to others, that the additional equation (I.13) requires only a particularization of the constant value of H.

Extremals of Zero-Length. In the particular case, previously excluded, $ds^2 = 0$, the equation

$$2\mathfrak{x} = 0$$

is by hypothesis satisfied along the corresponding class of solutions, and then we may no longer particularize the independent variable τ by a contradictory equation as $2\mathfrak{x}=1$. But conversely, leaving τ undetermined, we may proceed exactly as before, the only difference being that the constant 1 of (I.6) is here replaced by 0. The conclusion is that the canonical system (I.14) is still valid, the constant value of H being 0 instead of $mc/2$.

APPENDIX II

Illustrative Examples of Several Devices by Which Transformations in General Coordinates May be Performed. Let us consider, with reference to two independent variables x and y, the differential equations

$$\text{(II)} \qquad \partial\phi/\partial x - \partial\psi/\partial y = 0, \quad \partial\phi/\partial y + \partial\psi/\partial x = 0,$$

which express conditions that $\phi+i\psi$ be a (monogenic) function of the complex variable $x+iy$.

How are these equations to be transformed, in passing to new variables x^1, x^2, bound to x, y by given relations?

If no additional condition is added, the question admits, as we shall see, of several solutions, all substantially equivalent, but leading to different forms of the equations referred to new variables.

In order to obtain a clear understanding in this regard, let us adopt successively three criteria, all reasonable, for the transformation of (II).

(a) As an entirely suitable convention, we may *assume* that the functions ϕ and ψ behave both as *invariants*. This means that, when the x, y are replaced by the new variables x^1, x^2, the transformed ϕ, ψ are the same as before, with the understanding, however, that the x, y must be replaced by their expressions as functions of x^1, x^2.

Let us confine ourselves at first to *isometric* transformations, i.e. those arising from a unique relation $z'=f(z)$, where $z=x+iy$ as before, $z'=x^1+ix^2$, and f is an analytic function. Then the transformed $\phi(x^1, x^2)$, $i\psi(x^1, x^2)$ are still the real and imaginary parts of a function of the complex variable z', and the equations (II) become

(II.1) $\partial\phi/\partial x^1 - \partial\psi/\partial x^2 = 0, \quad \partial\phi/\partial x^2 + \partial\psi/\partial x^1 = 0,$

thus preserving the form (II).

(b) We pass now to general changes of variables (no longer isometric), still assuming for ϕ and ψ the law of invariance.

To get the equations corresponding to (II) in the new variables x^1, x^2, we have only to introduce, instead of $\partial/\partial x, \partial/\partial y$, the operators

$$(\partial x^1/\partial x)(\partial/\partial x^1) + (\partial x^2/\partial x)(\partial/\partial x^2),$$
$$(\partial x^1/\partial y)(\partial/\partial x^1) + (\partial x^2/\partial y)(\partial/\partial x^2).$$

The transformed equations then look rather complicated, in no way resembling the original equations. But it is possible to give them an invariant structure by the aid of geometric subsidiary elements. If we think of x, y as being cartesian coordinates, and introduce the square length

$$ds^2 = dx^2 + dy^2,$$

we shall have, in general coordinates x^1, x^2,

(II.2) $$ds^2 = a_{\mu\nu}dx^\mu dx^\nu,$$

the coefficients $a_{\mu\nu}(\mu, \nu = 1, 2)$ thus providing the metric tensor.

The cartesian lattice $x = \text{constant}$, $y = \text{constant}$ is formed by two congruences of (straight) lines, each of which has a proper geometrical meaning. The partial derivatives $\partial/\partial x, \partial/\partial y$ may be regarded as directional derivatives $d/ds_1, d/ds_2$ along the lines $y = \text{constant}$, and $x = \text{constant}$, respectively. They obtain in this way an invariant determination which may be easily put in explicit form with respect to our generic (curvilinear) coordinates x^1, x^2. We need only introduce, with reference to these coordinates, the parameters $\lambda_1{}^\mu, \lambda_2{}^\mu (\mu = 1, 2)$ of the lines $y = \text{constant}$, $x = \text{constant}$, respectively. We have then

$$\partial/\partial x = d/ds_1 = \lambda_1{}^\mu \partial/\partial x^\mu, \quad \partial/\partial y = d/ds_2 = \lambda_2{}^\mu \partial/\partial x^\mu,$$

and (II.1) takes the form

(II.3) $$d\phi/ds_1 - d\psi/ds_2 = 0, \quad d\phi/ds_2 + d\psi/ds_1 = 0,$$

which is, in the explained sense, completely invariant.

In the particular case of an isometric transformation, these equations may be linearly combined so as to assume the previous form (II.1).

(c) Instead of regarding ϕ and ψ as invariant functions, it is possible to make other assumptions; we may for instance suppose that, when variables are changed, not the single ϕ and ψ, but the differential form

$$\phi dx - \psi dy$$

has invariant character. This means that, if in general coordinates x^1, x^2 the linear form $\phi dx - \psi dy$ becomes

$$v_1 dx^1 + v_2 dx^2,$$

the coefficients v_1, v_2 are to be considered as the *transforms* of ϕ and ψ (law of *covariance*). The explicit formulas are

$$\begin{aligned} v_1 &= \phi \partial x/\partial x^1 - \psi \partial y/\partial x^1, \\ v_2 &= \phi \partial x/\partial x^2 - \psi \partial y/\partial x^2. \end{aligned} \tag{II.4}$$

On this assumption v_1, v_2 may be interpreted as *covariant components* of a vector v, precisely of that vector which had, in the original cartesian coordinates, the components ϕ and $-\psi$.

Let us now proceed to the transformation of the differential system (II) by introduction into it of new variables (x^1, x^2 instead of x, y) and also new functions (v_1, v_2 instead of ϕ, $-\psi$).

The second equation of (II) is simply the condition for the Pfaffian

$$\phi dx - \psi dy = v_1 dx^1 + v_2 dx^2$$

to be an exact differential. As this property always belongs to the same Pfaffian, whatever variables are adopted, we must have

$$\partial v_1/\partial x^2 - \partial v_2/\partial x^1 = 0, \tag{II.5}$$

which supplies one of the transformed equations. In vectorial language it expresses the fact that the vector-field v is a *conservative one*, or that a function V of position exists, whose gradient is v.

To complete without direct calculation the transformation of (II), it will be useful, here again, to appeal to the euclidean metric, defined by

$$ds^2 = dx^2 + dy^2 = a_{\mu\nu} dx^\mu dx^\nu.$$

On account of this metric we may introduce, at any point x, y, besides v, a second vector w through a rotation of 90° (in the positive sense of the cartesian lattice, Oxy). The cartesian components of w are then ψ, ϕ; and the first of the equations (II) is only the expression of the fact that the vector-field w admits a potential function. It may therefore be transformed, as was the other, in general coordinates x^1, x^2, giving

$$\partial w_1/\partial x^2 - \partial w_2/\partial x^1 = 0, \tag{II.6}$$

where w_1, w_2 denote clearly the covariant components of w to be calculated from (II.3) (by writing there ψ and ϕ instead of ϕ and $-\psi$).

Synthetically we may conclude that, by the introduction of euclidean metric and, at every point, of a second vector w (orthogonal to v and of equal length), both equations (II) have invariant interpretation.

They are merely the conditions that the fields v and w each admit a potential function, that is to say are irrotational, or finally that they have vanishing curl round any circuit (reducible to a point, within the field).

(d) But we may go a step further and reach another invariant interpretation, which requires only one of the two vector-fields v and w: say v. In order to see this we need only to remember that w is perpendicular to v and has equal length. Hence, along a generic curve s, if ds and dn denote the elements of arc and of the normal (oriented as dx, dy), we have for the components of v, w,

$$v_s = w_n, \qquad v_n = -w_s,$$

and therefore, for any arc s,

$$\int_s v_s ds = \int_s w_n ds, \qquad \int_s v_n ds = -\int_s w_n ds,$$

so that the vanishing of the *curl* of one vector, around any closed line, implies the vanishing of the *flow* of the other vector, and conversely. We may consequently replace the statement under (c), equivalent to (II), by the following:

In the vector-field v, curl and flow are both zero (i.e., the field is at the same time conservative and solenoidal).

In the usual vector symbolism we may write

$$\text{(II.7)} \qquad \text{rot } v = 0, \qquad \text{div } v = 0,$$

or explicitly

$$\text{(II.8)} \qquad v_{12} - v_{21} = 0, \qquad a^{\mu\nu} v_{\mu\nu} = 0,$$

where $v_{\mu\nu}$ are covariant derivatives of v_μ with respect to the euclidean form (II.2) and $a^{\mu\nu}$ coefficients of its reciprocal. The first of these equations is not only equivalent to, but even identical with (II.5), by a known relation between covariant and ordinary derivatives.

Apart from (II.1), which concern only isometric transformations, the other forms (II.3); (II.5); (II.6); (II.7); or (II.8), in which the original system (II) may be represented referred to general coordinates, are all substantially equivalent, in the sense that the two equations belonging to one of them are linear combinations of two others, with at most further linear substitution in the two dependent variables.

In the abstract the various systems arrived at are on the same footing. Only particular reasons, arising from the nature of things represented by the initial equations (II), or ease of calculation, may lead to the preference of one form to another.

Remark: To get from (II) any one of the transformed systems (II.3); (II.5); (II.6); (II.7); (II.8), in general coordinates, we have treated the original variables x, y as cartesian coordinates, employing the auxiliary $ds^2 = dx^2 + dy^2$. It is to be noted that every transformed system has a perfectly defined meaning even if the $ds^2 = a_{\mu\nu} dx^\mu dx^\nu$ involved is no longer euclidean. Then there do not exist special coordinates x, y for which the ds^2 takes the typical form $dx^2 + dy^2$, and therefore the general forms just quoted are no longer transforms of (II), though equivalent to each other. This happens often in the physical applications, where differential systems (like (II) but much more complex) are first established from premises connected with euclidean space, and then inferred for spaces of any metrical structure, through their transforms in general coordinates. Here we have Ricci's device for passing from euclidean to any metric. It rests on the belief that Riemann's curvature does not *directly* affect the natural laws; thus the passage to any metric becomes uniquely determined (at least for the more common theories).

APPENDIX III

Atomic Wave-Trains. Analysis and Physical Generalization of de Broglie's Inferences. Let us consider a space-time manifold V_4.

If τ represents any auxiliary independent variable, parametric equations

$$x^\mu = x^\mu(\tau), \qquad (\mu = 0, 1, 2, 3),$$

define a curve in V_4, or a motion of a point P. If, along this motion, $ds^2 > 0$, as for material bodies, it is convenient and customary to take at once $d\tau = ds$, the motion being thus referred to the proper time of P. But if $ds^2 = 0$, which means that the four-dimensional path is of zero-length (see end of Appendix I), it is meaningless to appeal to ds, and the introduction of some other parameter τ becomes unavoidable. At any rate the differential quotients of the x^μ's with respect to τ (which, for a proper motion, that is, for $ds^2 > 0$, are proportional to the dx^μ/ds) may be regarded as the contravariant components of a four-vector w.

Let us put

$$L(x\,|\,\dot{x}) = g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu/2, \qquad \text{(dot meaning } d/d\tau\text{)}, \tag{III.1}$$

and

$$p_\mu = \partial L/\partial\dot{x}^\mu = g_{\mu\nu}\dot{x}^\nu, \qquad (\mu = 0, 1, 2, 3), \tag{III.2}$$

obtaining in this way the covariant components of the same vector w.

Obviously we may regard w as the sum of two vectors having respectively $(p_0, 0, 0, 0)$ and $(0, p_1, p_2, p_3)$ as covariant components: w is then resolved into two vectors w_0 and w_1, the former perpendicular and the latter tangent to the spacial section $x^0 =$ constant, through P.†

We know from the restricted relativity, where $x^0 = ct$ and the p's are identical with the corresponding $\dot{x}$'s, that the ratio of the absolute values of w_0 and w_1,

† It would be equally permissible to decompose w, through its contravariant components, into the two vectors $(\dot{x}^0, 0, 0, 0)$, $(0, \dot{x}^1, \dot{x}^2, \dot{x}^3)$. Then the former has the direction of the time-like line $x^0(x^1 = \text{constant}, x^2 = \text{constant}, x^3 = \text{constant})$, the latter still being tangent to the spacial section. The two (covariant and contravariant) decompositions coincide if, and only if, the x^0-lines are orthogonal to the surfaces $x^0 = \text{constant}$, i.e., if $g_{01} = g_{02} = g_{03} = 0$.

$$w_0 = |p_0|, \qquad w_1 = |(p_1^2 + p_2^2 + p_3^2)^{1/2}|,$$

behaves as energy divided by c to momentum q of the moving point, i.e.,

$$w_0/w_1 = E/(cq). \tag{III.3}$$

For a general V_4 we are naturally led to assume that (III.3) still holds as a mechanical interpretation of the geometrical decomposition of w. On account of the fourfold metric

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu,$$

$g^{\mu\nu}$ being the reciprocal elements, we have for the merely temporal vector w_0, of covariant components $(p_0, 0, 0, 0)$, the square length

$$w_0 = g^{00}p_0^2. \tag{III.4}$$

For the merely spacial vector w_1 of covariant components $(0, p_1, p_2, p_3)$, we must appeal to the metric, subordinated by the former ds^2 in a section $x^0 = \text{constant}$, i.e., to

$$-dl^2 = \sum_1^3{}_{\mu\nu} g_{\mu\nu}dx^\mu dx^\nu. \tag{III.5}$$

It will accordingly be convenient to introduce, besides the $g^{\mu\nu}$, the coefficients $g'^{\mu\nu}$ of the *ternary* form reciprocal to dl^2 (which coincide with $-g^{\mu\nu}$ only when $g_{01} = g_{02} = g_{03} = 0$). The three-dimensional contravariant components p_μ' of w_1 are then

$$p'^\mu = \sum_1^3{}_\nu g'^{\mu\nu}p_\nu, \tag{III.6}$$

and therefore

$$w_1^2 = \sum_1^3{}_\mu p_\mu p'^\mu = \sum_1^3{}_{\mu\nu} g'^{\mu\nu}p_\mu p_\nu. \tag{III.7}$$

Now let us especially suppose that the p_μ verify equation (12) of the text, i.e.,

$$H(p \mid x) = g^{\mu\nu}p_\mu p_\nu/2 = 0, \tag{III.8}$$

which, owing to the identity

$$2L = g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = g^{\mu\nu}p_\mu p_\nu = 2H,$$

expresses the fact that the four-vector w is of zero length. Furthermore, if we treat the p_μ as derivatives of an unknown function $z(x^0, x^1, x^2, x^3)$, equation (III.8) may be solved with respect to p_0 and we get equation (13) of the text, that is,

$$p_0 + \mathfrak{H}(p_1, p_2, p_3 \mid x) = 0. \tag{III.9}$$

Let us fix our attention on any particular solution

$$z(x^0, x^1, x^2, x^3)$$

of (III.8). By virtue of the homogeneity, in z, any function $f(z)$ is also a solution.

In the neighborhood of any point P of V_4, that is, of any system of values x^μ, we have

$$dz = p_0 dx^0 + \sum_1^3{}_\mu p_\mu dx^\mu, \tag{III.10}$$

dx^0, dx^μ being arbitrary. Let us consider x^1, x^2, x^3 separately from x^0, and look provisionally at z as depending on x^1, x^2, x^3, while x^0 acts as parameter. Its derivatives $p_\mu (\mu = 1, 2, 3)$ are the covariant components of the vector w_1. We have therefore on one hand

$$w_1 = \text{grad}' z,$$

the accent alluding to a mere three-dimensional gradient; and besides, from (III.7),

$$\Delta' z = \sum_1^3{}_{\mu\nu} g'^{\mu\nu} p_\mu p_\nu = w_1^2, \tag{III.11}$$

$\Delta' z$ being Beltrami's parameter of the function z with respect to the spacial metric having dl^2 as square element of line.

Let us now resume using from (III.6) the contravariant components p'^μ of w_1 (in this metric). The parameters of the corresponding direction are

$$p'^\mu / w_1 = p'^\mu / (\Delta' z)^{1/2}.$$

The spacial displacement dx^μ in (III.10) may be resolved into two components: the one of algebraic value $-dn$, directed along $w_1 = \text{grad}' z$, and therefore of contravariant components

$$-dn(p'^\mu / w_1),$$

and the other perpendicular to w_1. This latter contributes nothing to the scalar product by w_1, i.e., to the sum $\sum_{\mu=1}^{3} p_\mu dx^\mu$. Therefore we may write, by virtue of (III.7), (III.4),

$$\sum_{1}^{3} p_\mu dx^\mu = - w_1 dn,$$

$$p_0 dx^0 = (w_0/(g^{00})^{1/2})dx^0,$$

so that, by (III.10), in the infinitesimal neighborhood of P, the expression of our integral z reduces to

$$z = z_p + (w_0/(g^{00})^{1/2})dx^0 - w_1 dn. \tag{III.12}$$

Hence we recognize that, in the immediate neighborhood of any instant x^0 and place x^1, x^2, x^3 (our point P of V_4), z *may be regarded as a linear function of only two independent arguments*: *the one*, dx^0, *or* dx^0/c, *expressing* (this latter ordinary) *time, and the other*, dn, *expressing length in the spacial direction of* grad′z.

In the same infinitesimal range, with

$$z_p + (cw_0/(g^{00})^{1/2})dx^0/c - w_1 dn,$$

any function whatever of this argument is again, as observed just above, a solution of the same homogeneous equation (III.8) or (III.9).

We may assume in particular as a solution

$$z = ae^{2\pi i\alpha\{(cw_0/(g^{00})^{1/2})dx^0/c - w_1 dn\}},$$

where

$$a = e^{2\pi i z_P}$$

and α is an arbitrary constant (absolute, or depending upon P). This *wave-function* z in this form is familiar as the ordinary representation of vibratory progressive waves; therefore

$$\alpha c w_0/(g^{00})^{1/2}$$

is to be interpreted as *instantaneous frequency* ν, and $1/(\alpha w_1)$ as *local wave-length* λ. At any rate the *velocity of propagation* V has the value

$$V = \lambda\nu = (cw_0)/((g^{00})^{1/2}w_1) \tag{III.13}$$

independent of the arbitrary constant α.

Now one can associate with wave-trains of this kind the corresponding bicharacteristics (15). The $p_\mu(\mu=0, 1, 2, 3)$ just dealt with are then to be interpreted as covariant components of a fourfold vector w, related by (a), (b), (c) of § 16 to the motion of particles of mass m. We have accordingly from (III.3) the corpuscular interpretation of w_0/w_1, $E/(mcv)$, where v and E represent their (mean) individual velocity and energy.

Introducing this value of w_0/w_1 in (III.13), we have

$$\lambda\nu = E/((g^{00})^{1/2}mv),$$

and finally, if ν and E are correlated by Planck's law $E=h\nu$,

$$\lambda = h/((g^{00})^{1/2}mv), \tag{III.14}$$

which is the expression of de Broglie's wave-length, extended from pseudo-euclidean to any space-time manifold, with explicit statement of the physical assumptions on which it rests.

THE UNIVERSITY OF ROME

MATRICES WITH ELEMENTS IN A PRINCIPAL IDEAL RING*

BY C. C. MACDUFFEE

1. *Rings.* To attempt to distinguish between algebra and number theory is probably futile, but, speaking approximately, it may be said that algebra (in the narrowest sense of the word) is the study of fields, while number theory is the study of rings. The mathematical system which seems most satisfactory as an abstraction of the system of rational integers is the principal ideal ring. By this I mean that the basic theorems of number theory, such as unique factorization into primes, hold for a principal ideal ring, while the concept of principal ideal ring is sufficiently general to include many other instances besides the rational integers.

A *ring*† is a mathematical system composed of more than one element, an equals relation, and two operations, $+$ and $\times$, subject to the following laws. The elements form an abelian group relative to the operation $+$, the identity element being denoted by 0. The set of elements is closed under the operation $\times$, which is associative. Finally, the operation $\times$ is distributive with respect to the operation $+$.

If $a \neq 0$ and $b \neq 0$ are elements of a ring $\mathfrak{R}$ such that $ab = 0$, then a and b are called *divisors of zero.* A commutative ring without divisors of zero is called a *domain of integrity.*

Let a, b, c be elements of a domain of integrity $\mathfrak{D}$. If $ab = c$, then $a \mid c$ (a divides c), $b \mid c$, and a and b are called *divisors* of c. If $a \mid b$ and $a \mid c$, then a is called a *common divisor* of b and c. If, furthermore, every common divisor of b and c divides a, then a is a *greatest common divisor* (g. c. d.) of b and c.

If there exists a number 1 of $\mathfrak{D}$ such that $1 \cdot a = a \cdot 1 = a$ for every a, this number 1 is called a *principal unit.*‡ A domain of integrity with a principal unit in which every pair of elements

* Symposium lecture delivered at the meeting of the Society in Chicago, April 15, 1933.

† In the interest of uniformity I have used the definitions of van der Waerden, *Moderne Algebra*, Springer, 1930–31.

‡ *Einselement*, van der Waerden.

not both 0 have a greatest common divisor representable linearly in terms of the elements is called a *principal ideal ring.*

In a principal ideal ring $\mathfrak{P}$, an element which divides 1 is called a *unit.* The relation $a = ub$, where u is a unit is reciprocal, and the two numbers a and b so related are called *associates.* A set of numbers of $\mathfrak{P}$ no two of which are associated but such that every number of $\mathfrak{P}$ is associated with one of them will be called a *complete set of non-associates* for $\mathfrak{P}$. Thus the positive integers and 0 constitute a complete set of non-associates in the ring of rational integers.

A field* is an instance of a principal ideal ring, but from the standpoint of number theory it is a trivial instance, since every element except 0 is a unit. The polynomial domain of a field, that is, the set of all polynomials in one indeterminate with coefficients in the field, is a non-trivial instance. The units of this polynomial domain are the elements of the field (0 excluded), and the primes are the irreducible polynomials. Maximal domains of algebraic fields of class number unity are also non-trivial instances.

According to the definition of principal ideal ring, every two numbers a and b of $\mathfrak{P}$ have a g. c. d. d expressible in the form

$$d = pa + qb,$$

where p and q are in $\mathfrak{P}$. The determination of p and q is for most rings a practical problem of considerable difficulty. In the ring of rational integers it is handled by means of the well known Euclid algorithm. For polynomial domains and a very few algebraic rings a Euclid algorithm has been developed.

The term *euclidean ring* has been applied to a principal ideal ring with a Euclid algorithm, and it has been considered that the separation of rings into euclidean and non-euclidean rings was a fundamental separation. Recent developments have tended to question this, and to indicate that the separation into rings of class number unity (principal ideal rings) and those of higher class number is of much more importance. It is true that the well known methods of proof of the fundamental theorems of number theory carry over only to euclidean rings, but the theorems themselves are generally true for principal ideal rings.†

* *Kommutativ Körper*, van der Waerden.

† Van der Waerden, loc. cit., vol. I, §17.

Just what algebraic rings of class number unity have euclidean algorithms is not known. Only five quadratic fields of negative discriminant have an algorithm based on decreasing norms, while ten of positive discriminant are known to have such. Perron* has recently suggested that all quadratic rings of positive discriminant and class number unity may have such an algorithm, but could not prove this. Then, too, the existence of an algorithm based on some other stathm† than the norm is a possibility. Very little is known of algebraic rings of degree higher than the second.

It is thus evident that the whole matter of euclidean rings is in disorder. For practical purposes, a euclidean ring is one for which some person has discovered a Euclid algorithm.

One of the purposes of this paper is to add support to the point of view that the principal ideal ring is the important concept rather than the euclidean ring by giving a unified account of an extension to principal ideal rings of the more important results in the theory of matrices with rational integral elements.

2. *Matric Rings.* Consider a mathematical system $\mathfrak{M}$ whose elements are the arrays

$$A = \begin{Vmatrix} a_{11} & a_{12} \cdots a_{1n} \\ a_{21} & a_{22} \cdots a_{2n} \\ \cdots & \cdots \\ a_{n1} & a_{n2} \cdots a_{nn} \end{Vmatrix} = (a_{rs}),$$

where the a_{ij} belong to a ring $\mathfrak{R}$. Two arrays $A=(a_{rs})$ and $B=(b_{rs})$ are called *equal* if $a_{rs}=b_{rs}$ for every r and s.

The operation of addition $(+)$ is defined by

$$A + B = (a_{rs} + b_{rs}).$$

Evidently the arrays form an abelian group with respect to addition, since the same is true of the elements of the ring $\mathfrak{R}$. The identity for addition, composed entirely of 0's, will be denoted by O.

---

* Mathematische Annalen, vol. 107 (1932), pp. 489–495.

† See J. H. M. Wedderburn, Journal für Mathematik, vol. 167 (1931), pp. 129–141.

The operation of multiplication is defined by

$$AB = \left(\sum_i a_{ri} b_{is} \right).$$

The product is unique, and $OA = AO = O$ for every A.

It is easily seen that multiplication in $\mathfrak{M}$ is associative, and distributive with respect to addition, since the same is true in $\mathfrak{R}$.

Thus the system $\mathfrak{M}$ is a ring, which we may call the *total matric ring of order n over* $\mathfrak{R}$. Each of the arrays composing $\mathfrak{M}$ will be called a *matrix*.

If the ring $\mathfrak{R}'$ has a principal unit 1, the derived matric ring $\mathfrak{M}'$ has the principal unit $I = (\delta_{rs})$, where δ_{rs} is 1 or 0 according as $r = s$ or $r \neq s$. We shall call I the *identity matrix* of $\mathfrak{M}'$.

Further specialization of $\mathfrak{R}'$ to a commutative ring, or to a ring without divisors of zero, does not carry with it the corresponding specialization of $\mathfrak{M}'$.

If $\mathfrak{R}'$ is a ring with a principal unit, the matrices $S(k) = (k\delta_{rs})$ constitute a subring of $\mathfrak{M}'$ which is isomorphic with the numbers k of $\mathfrak{R}'$. The operation of scalar multiplication by which a matrix A of $\mathfrak{M}'$ is multiplied by a number k of $\mathfrak{R}'$ consists in replacing k by $S(k)$ and forming the matric product $S(k)A$.

The matrix $A^{\mathsf{T}} = (a_{sr})$, obtained from $A = (a_{rs})$ by changing rows to columns, is called the *transpose* of A. A matrix S such that $S^{\mathsf{T}} = S$ is called *symmetric*. A matrix Q such that $Q^{\mathsf{T}} = -Q$ is called *skew*.

If A is a matrix of $\mathfrak{M}$, then $2A$ can be represented as a sum of the symmetric matrix $A + A^{\mathsf{T}}$ and the skew matrix $A - A^{\mathsf{T}}$. If 2 is not a divisor of zero in the ring $\mathfrak{R}$, this representation is unique. For if

$$2A = S + Q,$$

where S is symmetric and Q is skew, then

$$2A^{\mathsf{T}} = S - Q,$$

so that

$$2[A + A^{\mathsf{T}} - S] = 0, \quad 2[A - A^{\mathsf{T}} - Q] = 0.$$

3. *Unimodular Matrices.* Let us now consider the total matric ring $\mathfrak{M}$ of matrices with elements in a principal ideal ring $\mathfrak{P}$.

A matrix U of $\mathfrak{M}$ is called *unimodular*, or a *unit matrix*, if there exists a matrix U' such that $UU'=I$.

Denote the determinant of U by $d(U)$. Since $UU'=I$ implies $d(U)d(U')=1$, $d(U)$ is a unit of $\mathfrak{P}$. Conversely if U is in $\mathfrak{M}$ and $d(U)$ is a unit of $\mathfrak{P}$, then U^1 (the inverse of U) is in $\mathfrak{M}$, and serves as the U' of the definition. Hence also $U'U=I$, and U' is a unit.

Similarly a matrix A of $\mathfrak{M}$ is a divisor of zero if and only if $d(A)=0$.

THEOREM 1. *Let $a_1, a_2, \cdots, a_n$, all numbers of a principal ideal ring $\mathfrak{P}$, have the greatest common divisor d_n. There exists a matrix of determinant d_n having $a_1, a_2, \cdots, a_n$ as its first row.*

This theorem was first given for rational integers, $n=3$, by G. Eisenstein,* and for any n by C. Hermite.† The following proof (for the case of rational integers) was given by A. Bloch.‡

The theorem is evidently true for $n=2$, for if $pa_1+qa_2=d_2$, then

$$\begin{Vmatrix} a_1 & a_2 \\ -q & p \end{Vmatrix} = d_2.$$

Suppose that it holds for $n-1$, and let D_{n-1} be a matrix which has $a_1, a_2, \cdots, a_{n-1}$ as its first row, and whose determinant is the g. c. d. d_{n-1} of $a_1, a_2, \cdots, a_{n-1}$. Determine p and q so that $pd_{n-1}-qa_n=d_n$. Let

$$D_n = \begin{Vmatrix} & & & a_n \\ & D_{n-1} & & 0 \\ & & & \cdots \\ & & & 0 \\ \dfrac{a_1q}{d_{n-1}} & \dfrac{a_2q}{d_{n-1}} & \cdots \dfrac{a_{n-1}q}{d_{n-1}} & p \end{Vmatrix}.$$

Then, expanding according to the elements of the last column,

$$\begin{aligned} d(D_n) &= (-1)^{n-1}a_n \frac{q}{d_{n-1}}(-1)^n d_{n-1} + pd_{n-1} \\ &= pd_{n-1} - qa_n = d_n. \end{aligned}$$

* Journal für Mathematik, vol. 28 (1884), pp. 289–374.

† Journal de Mathématiques, (1), vol. 14 (1849), pp. 21–30.

‡ Bulletin de la Société Mathématique, vol. 50 (1922), pp. 100–110.

Other proofs have been given by K. Weihrauch,* Bianchi,† and H. Hancock.‡

4. *Types of Equality in Matric Theory.* At the very basis of every mathematical system lies the notion of equality. The abstract formulation of this notion may be embodied in the following four postulates.§ The relation $A=B$ is a relation of equality if it is

(1.) *Determinative* (Either $A=B$ or $A\neq B$),

(2.) *Reflexive* ($A=A$),

(3.) *Symmetric* ($A=B$ implies $B=A$),

(4.) *Transitive* (If $A=B$ and $B=C$, then $A=C$).

Such a definition of equality constitutes a separation of the elements into classes.

The richness of the matric theory is due in large part to the number of non-isomorphic types of equality which can be defined, each having associated with it an interesting and fairly extensive theory. A few examples follow.

(1.) Let $\mathfrak{R}'$ be a ring with a principal unit. If there exists a unimodular matrix U such that $A=UB$, then A is a *left associate* of B, written $A\overset{L}{=}B$.

(2.) If A is a *right associate* of B, then $A\overset{R}{=}B$.

(3.) If there exist two unimodular matrices U and V such that $A=UBV$, then $A\overset{E}{=}B$. (A is *equivalent* to B.)

(4.) If there exists a unimodular matrix U such that $A=U^{\mathsf{T}}BU$, then $A\overset{C}{=}B$. (A is *congruent* with B.)

(5.) If there exists a unimodular matrix U such that $A=U^{\mathsf{I}}BU$, then $A\overset{S}{=}B$. (A is *similar* to B.)

(6.) If there exists an orthogonal matrix U ($U^{\mathsf{I}}=U^{\mathsf{T}}$) such that $A=U^{\mathsf{T}}BU$, then $A\overset{O}{=}B$. (A is *orthogonally congruent* with B.)

All of these relationships obey the four postulates stated above, and are therefore in an abstract sense relations of equality. These examples by no means exhaust the possibilities.

---

* Zeitschrift für Mathematik und Physik, vol. 21 (1876), pp. 134–137.

† *Lezioni sulla Teoria dei Numeri Algebrici*, pp. 1–7.

‡ American Mathematical Monthly, vol. 31 (1924), pp. 161–162.

§ See O. Ore, this Bulletin, vol. 37 (1931), p. 538. Van der Waerden (vol. I, p. 14) would call this relation *equivalence*, and say that the class C_a containing a is *equal* to the class C_b containing b if and only if a is equivalent to b. It seems to the writer that Ore has carried this idea to its logical conclusion.

An interesting example of slightly more complicated nature was given by A. Loewy.* If the elements of A, B, and P are functions of x, and P^D is the matrix obtained from P by replacing each element by its derivative, and if

$$A = -P^D P^I + PBP^I,$$

then A is *similar to* B *in the sense of Loewy.*

5. *Associated Matrices.* The following *elementary operations* upon the rows of a matrix can be accomplished by multiplying the given matrix on the left by an *elementary matrix*, namely the unimodular matrix obtained by performing the desired elementary operation upon the identity matrix I.

(1.) The interchange of two rows.

(2.) The multiplication of the elements of a row by a unit u of $\mathfrak{P}$.

(3.) The addition to the elements of a row of k times the corresponding elements of another row, k being in $\mathfrak{P}$.

Every elementary matrix is unimodular, and its inverse is an elementary matrix of the same type. The theory of elementary matrices is due to L. Kronecker.†

If $\mathfrak{P}$ has a euclidean algorithm, every unimodular matrix is a product of elementary matrices. This appears to be not so in other rings.‡

If $UA = B$, then for every k

$$\sum u_{ij} a_{jk} = b_{ik},$$

so that every g. c. d. of the elements of the kth column of A is a common divisor of the elements of the kth column of B. If U is unimodular, the relation of A to B is reciprocal. Hence the g. c. d. of the elements of every column is invariant under transformations of this type.

These n invariants do not form a complete system, however, as the next theorem shows.

THEOREM 2. *Every matrix A with elements in $\mathfrak{P}$ is the left associate of a matrix having 0's above the main diagonal, each di-*

* Mathematische Annalen, vol. 78 (1918), pp. 1–51.

† Berliner Akademie, Monatsberichte, 1866, pp. 597–612.

‡ See van der Waerden, loc. cit., vol. II, p. 122.

agonal element lying in a prescribed system of non-associates, and each element below the main diagonal lying in a prescribed residue system modulo the diagonal element above it. If, furthermore, any diagonal element is 0, *all the elements of its row can be made* 0. *This form is unique.*

This theorem was stated for a non-singular matrix with rational integral elements by C. Hermite.*

To avoid being tiresome, I shall prove this merely for a special case. It is fairly evident that the procedure is general. Let

$$A = \begin{Vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{Vmatrix}.$$

Unless every element of the last column is zero, they have a g. c. d.

$$d_3 = b_1a_{13} + b_2a_{23} + b_3a_{33}$$

lying in any prescribed system of non-associates. By Theorem 1 there is a unimodular matrix U having b_1, b_2, b_3 as its last row. Then UA has d_3 in the (3, 3)-position, and every other element of the last column is a linear combination of a_{13}, a_{23}, a_{33}, and hence a multiple of d_3. By subtracting a proper multiple of the last row from each of the other rows, a matrix is obtained whose last column consists entirely of 0's above the main diagonal. Similarly, if we work now only with the second order minor in the upper left corner, a_{12} can be made 0.

In case every element of the last column of A is 0, the procedure must be slightly modified. Let

$$d_2 = b_1a_{12} + b_2a_{22} + b_3a_{32}$$

be the desired g. c. d. of the elements of the second column. Let U be unimodular with b_1, b_2, b_3 as its second row. Then UA still has all zeros in the third column, while d_2 is in the (2, 2)-position, and d_2 divides every other element of the second column. Thus A is the left associate of

---

* Journal für Mathematik, vol. 41 (1851), pp. 191–216.

$$\left\| \begin{matrix} a_{11} & 0 & 0 \\ a_{21} & d_2 & 0 \\ a_{31} & 0 & 0 \end{matrix} \right\|.$$

Let $d_1 = b_1 a_{11} + b_3 a_{31}$ be a g. c. d. of a_{11} and a_{31}. Take U unimodular so that

$$U = \left\| \begin{matrix} b_1 & 0 & b_3 \\ 0 & 1 & 0 \\ u_{31} & 0 & u_{33} \end{matrix} \right\|.$$

Then UA has d_1 in the (1, 1)-position, and d_1 is a divisor of the other elements of the first column, while the last two columns are intact. Thus A is the left associate of

$$\left\| \begin{matrix} d_1 & 0 & 0 \\ a_{21} & d_2 & 0 \\ 0 & 0 & 0 \end{matrix} \right\|.$$

Let us prove the uniqueness of this form for the example

$$A = \left\| \begin{matrix} a_{11} & 0 & 0 \\ 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} \end{matrix} \right\|, \quad a_{11} a_{33} \neq 0.$$

Suppose

$$\left\| \begin{matrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{matrix} \right\| \cdot \left\| \begin{matrix} a_{11} & 0 & 0 \\ 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} \end{matrix} \right\| = \left\| \begin{matrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{matrix} \right\|,$$

where $b_{ii} = 0$ implies $b_{ki} = 0$, $k = 1, 2, 3$ and (l_{rs}) is unimodular. Suppose that b_{ii} belongs to the same system of non-associates as a_{ii}, and that if $b_{ii} = a_{ii}$, then b_{ki} belongs to the same residue system modulo a_{ii} that a_{ki} does.

Since $l_{13} a_{33} = b_{13} = 0$ and $l_{23} a_{33} = b_{23} = 0$, and $a_{33} \neq 0$, then $l_{13} = l_{23} = 0$. Since (l_{rs}) is unimodular, l_{33} is a unit. Since a_{33} and $l_{33}\ a_{33} = b_{33}$ belong to the same system of non-associates, $l_{33} = 1$ and $b_{33} = a_{33}$.

Now $b_{22} = 0$, since $l_{23} = 0$. Hence $b_{21} = 0$ also, from the definition of canonical form. Hence $l_{21} a_{11} = b_{21} = 0$, $l_{21} = 0$. Then l_{11} and l_{22}

are units of $\mathfrak{P}$. But a_{11} and $l_{11}a_{11}=b_{11}$ are in the same system of non-associates, so $l_{11}=1$ and $b_{11}=a_{11}$. Since

$$l_{31}a_{11}+a_{31}=b_{31},$$

and a_{31} and b_{31} lie in the same residue system modulo $a_{11}=b_{11}$, it follows that $l_{31}=0$. Hence

$$(l_{rs})=\begin{Vmatrix} 1 & l_{12} & 0 \\ 0 & u & 0 \\ 0 & l_{31} & 1 \end{Vmatrix}.$$

This matrix leaves A unaltered when used as a left factor.

6. *Greatest Common Divisors.* If A, C, and D are matrices with elements in a principal ideal ring such that $A=CD$, then D is called a *right divisor* of A, and A is a *left multiple* of D. A *greatest common right divisor* (g. c. r. d.) D of two matrices A and B is a common right divisor which is a left multiple of every common right divisor of A and B. A *least common left multiple* (l. c. l. m.) of two matrices A and B is a common left multiple which is a right divisor of every common left multiple of A and B.

THEOREM 3. *Every pair of matrices A and B with elements in a principal ideal ring have a g. c. r. d. D expressible in the form*

$$D=PA+QB.$$

This theorem was first given for matrices A and B not both singular and with rational integral elements by du Pasquier* by means of a generalized Euclid algorithm. Computation by this method is laborious. The presentation here given is not only simple and independent of the Euclid algorithm (and hence extensible to principal ideal rings) but computation by means of it is easy and rapid. It is due in essence to E. Cahen,† and in the form here presented (for rational integers) to A. Chatelet.‡

By Theorem 2 there exists a unimodular matrix U such that

$$U\begin{Vmatrix} A & O \\ B & O \end{Vmatrix}=\begin{Vmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{Vmatrix}\begin{Vmatrix} A & O \\ B & O \end{Vmatrix}=\begin{Vmatrix} D & O \\ O & O \end{Vmatrix}.$$

* Naturforschende Gesellschaft zu Zürich, vol. 51 (1906), pp. 55–129.

† *Théorie des Nombres*, vol. I, 1914.

‡ *Groupes Abéliens Finis*, 1924,

Thus $U_{11}A+U_{12}B=D$, so that every common right divisor of A and B is a right divisor of D. Since U is unimodular, $U^1=V$ has elements in $\mathfrak{P}$, and

$$V\begin{Vmatrix} D & O \\ O & O \end{Vmatrix} = \begin{Vmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{Vmatrix} \begin{Vmatrix} D & O \\ O & O \end{Vmatrix} = \begin{Vmatrix} A & O \\ B & O \end{Vmatrix}.$$

Thus $A=V_{11}D$, $B=V_{21}D$, and D is a common right divisor of A and B.

If in the $2n\times n$ array $\binom{A}{B}$ is of rank n, the matrices A and B have a non-singular g. c. r. d. D. In this case every g. c. r. d. of A and B is a left associate of D. For if $D=PD_1$ and $D_1=QD$, $D=PQD$. If $d(D)\neq 0$, then $I=PQ$, whence P and Q are unimodular.

The above algorithm also furnishes a least common left multiple of A and B if both are non-singular. Evidently

$$U_{21}A + U_{22}B = O.$$

Let us define M by the formula

$$M = U_{21}A = -U_{22}B.$$

That is, M is a common left multiple of A and B. If M_1 is another c. l. m., their g. c. r. d. $M_2=PM+QM_1$ is a c. l. m. such that $M=HM_2$. Suppose $M_2=KA=LB$. Then

$$U_{21}A = HKA, \quad -U_{22}B = HLB,$$

and since A and B are non-singular,

$$U_{21} = HK, \; U_{22} = -HL.$$

But

$$I = U_{21}V_{12} + U_{22}V_{22} = H[HV_{12} - LV_{22}],$$

so H is unimodular, and M is a right divisor of M_1.

7. *Equivalence.* Two matrices A and B with elements in a principal ideal ring $\mathfrak{P}$ are *equivalent* if there exist two unimodular matrices U and V with elements in $\mathfrak{P}$ such that $A=UBV$.

With some modification of proofs, the theory of invariant factors and elementary divisors in an ordinary polynomial ring goes over intact to a principal ideal ring. First of all, we may note that each greatest common divisor d_i of the i-rowed minor

determinants of A is associated with every greatest common divisor d_i' of the i-rowed minor determinants of B. This follows from the fact that if $A = UBV$, every i-rowed minor of A can be written as a linear combination of the i-rowed minors of B, and if U and V are unimodular, the relation is reciprocal.

THEOREM 4. *Every matrix A of rank ρ with elements in $\mathfrak{P}$ is equivalent to a diagonal matrix $h_1, h_2, \cdots, h_\rho, 0, \cdots, 0$ where $h_i \neq 0$ and $h_i \mid h_{i+1}$.*

This theorem was given for rational integers by H. J. S. Smith.*

Since the Euclid algorithm is no longer available, Theorem 1 must be used as a substitute. After shifting rows and columns so that the minor of order ρ in the upper left corner is not 0, the element in the (1, 1)-position can be made $\neq 0$ and a g. c. d. of the elements of the first column, as was shown in the proof of Theorem 2. As in the usual reduction, the element in the (1, 1)-position either divides every remaining element of the first row, or it can be replaced by a proper divisor of itself—that is, replaced by a number having fewer prime factors. Since a number of $\mathfrak{P}$ neither 0 nor a unit has but a finite number of prime factors, this process can be repeated until a_{11} divides every other element of the first row and first column.† Now, as usual, A can be reduced by a repetition of this process to a diagonal form $[g_1, g_2, \cdots, g_\rho, 0, \cdots, 0]$.

By adding column 2, column 3, $\cdots$, column ρ to column 1, the matrix is made to assume the form

$$\left\| \begin{matrix} g_1 & 0 & 0 \cdots 0 \\ g_2 & g_2 & 0 \cdots 0 \\ g_3 & 0 & g_3 \cdots 0 \\ \cdots & \cdots & \cdots \end{matrix} \right\|.$$

As in the proof of Theorem 2, there is a unimodular matrix U which, used as a left factor, replaces g_1 by the g. c. d. h_1 of $g_1, g_2, \cdots, g_\rho$. Every element of the new matrix UA is a homo-

* Philosophical Transactions of the Royal Society of London, vol. 151 (1861–62), p. 314.

† Van der Waerden, loc. cit., vol. II, p. 124.

geneous linear combination of $g_1, g_2, \cdots, g_\rho$, so every element is divisible by h_1. Reduce every element of the first row and column except the element h_1 in the (1, 1)-position to 0. Continue the entire process with the $(n-1)$-rowed minor in the lower right corner to obtain an h_2 in the (2, 2)-position which divides $h_3, \cdots, h_\rho$. Finally we obtain the normal form.

Let $H = [h_1, \cdots, h_\rho, 0, \cdots, 0]$ be the normal form of A. The g. c. d. d_i of the i-rowed minors of A is associated with the g. c. d. d_i' of the i-rowed minors of H. Since $h_i \mid h_{i+1}$, it follows that $d_i = h_1 h_2 \cdots h_i$ divides $d_{i+1} = h_1 h_2 \cdots h_i h_{i+1}$, the quotient being h_{i+1}. Hence these quotients are invariants (up to a unit factor in $\mathfrak{P}$), so that the normal form of A is unique when we require that the h's shall lie in a specified system of non-associates.

In a principal ideal ring every element neither 0 nor a unit can be factored uniquely (except for unit factors) into a product of powers of primes.* Suppose

$$h_i = p_1^{e_{i1}} p_2^{e_{i2}} \cdots p_k^{e_{ik}}.$$

Since $h_i \mid h_{i+1}$, the exponents of each prime factor form a sequence

$$e_{nl} \geqq e_{n-1,l} \geqq \cdots \geqq e_{1l} \quad (l = 1, 2, \cdots, k).$$

Such of these powers $P^{e_{il}}$ as are not units are called the *elementary divisors* of A. They are defined up to unit factors.

This treatment unifies the elementary divisor theory of matrices with rational integral elements and that of the so-called λ-*matrices*, that is, matrices with elements in the polynomial domain of a field. Every polynomial domain (with one indeterminate) of a field is a principal ideal ring,† the units being the numbers $\neq 0$ of the field.

The theory of equivalent matrices is an excellent example of the interdependence of algebra and number theory. It seems to the writer that the elementary divisor theory appears simpler and more natural when it is looked upon as being fundamentally a branch of the theory of numbers.

8. *Congruence.* If there exists a unimodular matrix U such

* Van der Waerden, loc. cit., vol. I, p. 65.

† Van der Waerden, loc. cit., vol. I, p. 60.

that $A = U^T B U$, all matrices having elements in $\mathfrak{P}$, then A is said to be *congruent* with B.

Symmetric and skew matrices play an important role in the theory of congruence. Each of these properties is preserved under this relation. If $S+Q$ is congruent with S_1+Q_1, where S and S_1 are symmetric and Q and Q_1 are skew, then S is congruent with S_1 and Q with Q_1. Hence if $2A = S+Q$, the congruence invariants of S and Q are congruence invariants of A, and together they form a complete system for A except for rings in which 2 is a divisor of zero.

It is well known that the problem of the congruence of symmetric matrices is a problem of extreme difficulty. It is essentially the problem of the equivalence of quadratic forms in the theory of numbers. How refractory this problem is may be seen by glancing through Professor Dickson's recent book.*

It is perhaps less well known that the corresponding problem for skew matrices is capable of complete and simple solution not only for matrices with rational integral elements, but for matrices with elements in a principal ideal ring. It seems perverseness on the part of nature that an important problem should appear to be insoluble, while a similar problem of no evident importance should behave so well. Perhaps the ultimate explanation will consist in showing that the skew matrix is really important.† At present it is sufficient for the pure mathematician that the theory is elegant.

I wish at this point to call attention to a very useful notation due to A. Hurwitz.‡ If A and B are matrices, and O is a block of zeros, then

$$\left\| \begin{matrix} A & O \\ O & B \end{matrix} \right\|$$

is called the *direct sum* of A and B, and is written $A \dot{+} B$.

THEOREM 5. *Every skew matrix Q is of even rank $\rho = 2\mu$, and is congruent with a matrix*

---

* L. E. Dickson, *Studies in the Theory of Numbers*, University of Chicago Press, 1930.

† Professor Albert has called my attention to the fact that skew matrices are of prime importance in the theory of pure Riemann matrices.

‡ H. Kries, *Contribution à la Théorie des Systèmes Linéaires*, Zurich, 1906.

$$H = \left\| \begin{matrix} 0 & h_1 \\ -h_1 & 0 \end{matrix} \right\| \dot{+} \left\| \begin{matrix} 0 & h_2 \\ -h_2 & 0 \end{matrix} \right\| \dot{+} \cdots \dot{+} \left\| \begin{matrix} 0 & h_\mu \\ -h_\mu & 0 \end{matrix} \right\| \dot{+} \left\| \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \right\| \dot{+} \cdots,$$

where $h_i \mid h_{i+1}$. *The numbers* $h_1, h_1, h_2, h_2, \cdots, h_\mu, h_\mu$ *of the canonical form* H *are the invariant factors of* Q.

This theorem for the ring of rational integers was given by E. Cahen.*

I shall not give the details of this proof, for it follows the lines of the proofs of Theorems 2 and 4. The 0's in the main diagonal of the skew matrix enable one to work successively on rows and columns. In the case of symmetric matrices the non-zero elements in the main diagonal make a similar reduction impossible.

9. *Moduls.* The set $\mathfrak{L}$ of all numbers of the form

$$a_1\epsilon_1 + a_2\epsilon_2 + \cdots + a_n\epsilon_n,$$

where the a's range over a ring $\mathfrak{R}$ and the ϵ's are linearly independent with respect to $\mathfrak{R}$, constitute a *linear form modul.* The ϵ's constitute a *basis* for $\mathfrak{L}$.

If $U = (u_{rs})$ is unimodular with elements in $\mathfrak{R}$, then

$$\epsilon_i' = \sum u_{ij}\epsilon_j, \qquad (i = 1, 2, \cdots, n),$$

also constitute a basis for $\mathfrak{L}$, since every linear combination of the ϵ's is a linear combination of the ϵ''s, and vice versa. Conversely, every two bases of a linear form modul are so related by a unimodular transformation.

The theory of moduls over the ring of rational integers was developed along the lines which we shall follow by A. Chatelet.† The establishment of Theorems 1, 2, and 3 of this paper shows that Chatelet's results are true for every principal ideal ring.

---

* *Théorie des Nombres*, vol. I, 1914, p. 282.

† Annales de l'École Normale, (3), vol. 28 (1911), pp. 105–202; Comptes Rendus, vol. 154 (1912), p. 502; *Leçons sur la Thèorie des Nombres*, 1913; *Groupes Abéliens Finis*, 1924.

Let $\mathfrak{L}_1$ be a linear form sub-modul* of order n of $\mathfrak{L}$, and let $\mathfrak{L}_1$ have the basis $\lambda_1, \lambda_2, \cdots, \lambda_n$. Then

$$\lambda_i = \sum g_{ij}\epsilon_j, \qquad (i = 1, 2, \cdots, n),$$

where $G_1 = (g_{rs})$ is a non-singular matrix with elements in $\mathfrak{R}$. We may say that G_1 is *associated* with the basis $\lambda_1, \lambda_2, \cdots, \lambda_n$ of $\mathfrak{L}_1$. Every non-singular matrix G_1 determines in this way a basis of some linear form sub-modul of order n of $\mathfrak{L}$.

If $\mathfrak{L}_2$, with basis $\mu_1, \mu_2, \cdots, \mu_n$, is a linear form sub-modul of order n of $\mathfrak{L}_1$, every number of $\mathfrak{L}_2$ is in $\mathfrak{L}_1$, and in particular

$$\mu_i = \sum c_{ij}\lambda_j = \sum c_{ij}g_{jk}\epsilon_k, \quad (i = 1, 2, \cdots, n).$$

The matrix G_2 associated with the basis $\mu_1, \mu_2, \cdots, \mu_n$ of $\mathfrak{L}_2$ is CG_1, where C is a non-singular matrix with elements in $\mathfrak{R}$. Thus $\mathfrak{L}_1$ contains $\mathfrak{L}_2$ if and only if G_1 is a right divisor of G_2.

In particular, two moduls $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are equal if and only if G_1 and G_2 are left associates. A unique canonical basis for $\mathfrak{L}_1$ may be determined in accordance with Theorem 1.

Now suppose that $\mathfrak{R}$ is a principal ideal ring $\mathfrak{P}$. The set of numbers common to two moduls $\mathfrak{L}_1$ and $\mathfrak{L}_2$ constitute a modul $\mathfrak{L}_d$ called the *greatest common sub-modul* of $\mathfrak{L}_1$ and $\mathfrak{L}_2$. It may also be defined as that common sub-modul of $\mathfrak{L}_1$ and $\mathfrak{L}_2$ which contains every common sub-modul of $\mathfrak{L}_1$ and $\mathfrak{L}_2$. If G_1, G_2, and G_d are matrices associated with $\mathfrak{L}_1$, $\mathfrak{L}_2$, and $\mathfrak{L}_d$ respectively, then G_d is a right divisor of G_1 and G_2, and every common right divisor of G_1 and G_2 is a right divisor of G_d. That is, G_d is a g. c. r. d. of G_1 and G_2. Now G_d is unique up to a unimodular left factor—the same latitude of definition as the g. c. r. d. of two matrices.

In a similar manner all numbers contained in either $\mathfrak{L}_1$, or $\mathfrak{L}_2$, together with their sums and differences, constitute a modul $\mathfrak{L}_m$ called the *least common super-modul* of $\mathfrak{L}_1$ and $\mathfrak{L}_2$. It may also be defined as that modul containing $\mathfrak{L}_1$ and $\mathfrak{L}_2$ which is contained in every modul containing $\mathfrak{L}_1$ and $\mathfrak{L}_2$. If G_m is a matrix associated with $\mathfrak{L}_m$, it is clear from this last definition that G_m is a l. c. l. m. of G_1 and G_2.

10. *Ideals.* Let $\mathfrak{S}$ be a linear form modul, with coefficients in a principal ideal ring $\mathfrak{P}$, which is also a ring, and whose elements

* If $\mathfrak{R}$ is a principal ideal ring, every sub-modul of $\mathfrak{L}$ is a linear form modul. Van der Waerden, loc. cit., vol. II, p. 121.

are commutative with those of $\mathfrak{P}^*$. If $\mathfrak{S}$ has the basis $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$, then

$$\epsilon_i\epsilon_j = \sum_k c_{ijk}\epsilon_k,$$

where the constants of multiplication c_{ijk} are in $\mathfrak{P}$. Define the matrices

$$R_i = (c_{isr}), \qquad S_i = (c_{ris}).$$

Then both of the matric rings

$$R(\xi) = x_1R_1 + x_2R_2 + \cdots + x_nR_n,$$
$$S(\xi) = x_1S_1 + x_2S_2 + \cdots + x_nS_n$$

are isomorphic under both addition and multiplication with the numbers

$$\xi = x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_n\epsilon_n$$

of $\mathfrak{S}$. Unless $\mathfrak{S}$ is of very special form† the R_i are linearly independent, and the S_i are also. Thus we have two representations of $\mathfrak{S}$ as a matric ring.

A sub-modul of $\mathfrak{S}$ which is closed under multiplication on the left by numbers of $\mathfrak{S}$ is called a *left ideal*.‡ Similarly *right ideals* and *two-sided ideals* may be defined.

It is known that every ideal has a *minimal basis* $\lambda_1, \lambda_2, \cdots, \lambda_n$ such that every number of the ideal is given by the form

$$h_1\lambda_1 + h_2\lambda_2 + \cdots + h_n\lambda_n,$$

where the h's are in $\mathfrak{P}$; and if the ideal does not consist exclusively of divisors of zero, the representation is unique, that is, the λ's are linearly independent with respect to $\mathfrak{P}$.§

If $\lambda_1, \lambda_2, \cdots, \lambda_n$ form a minimal basis for a left ideal $\mathfrak{J}$, where

$$\lambda_i = \sum g_{ij}\epsilon_j,$$

every number κ of $\mathfrak{J}$ is of the form

---

* An instance of such a system is a domain of integrity of a linear associative algebra in the sense of Dickson. *Algebren und ihre Zahlentheorie*, 1927, p. 154.

† This Bulletin, vol. 35 (1929), pp. 344–349.

‡ Van der Waerden, loc. cit., vol. I, p. 53.

§ Transactions of this Society, vol. 31 (1929), p. 74.

$$\kappa = k_i\lambda_i = \sum k_i g_{ij}\epsilon_j,$$

while every number σ of $\mathfrak{S}$ is of the form

$$\sigma = \sum s_l\epsilon_l.$$

Since $\sigma\kappa$ is in $\mathfrak{J}$ for all values of s_l and k_i, there exist numbers d_r of $\mathfrak{P}$ such that

$$\sigma\kappa = \sum s_l k_i g_{ij} c_{ljh}\epsilon_h = \sum d_r g_{rt}\epsilon_t.$$

Hence

$$\sum s_l k_i g_{ij} c_{ljh} = \sum d_r g_{rh}.$$

In particular, when $s_l = \delta_{lp}$ and $k_i = \delta_{iq}$, there exist values d_{pqr} of d_r. For these values

$$\sum g_{qj} c_{pjh} = \sum d_{pqr} g_{rh},$$

or

$$GR_p^{\mathsf{T}} = D_p G, \quad D_p = (d_{prs}), \quad (p = 1, 2, \cdots, n). \tag{1}$$

That is, if the modul with which G is associated is a left ideal, matrices $D_1, D_2, \cdots, D_n$ with elements in $\mathfrak{P}$ exist satisfying the above condition.

The existence of the D's in (1) is a sufficient as well as a necessary condition that G be associated with an ideal. Let d_{pqr} and g_{qj} be numbers of $\mathfrak{P}$ satisfying the above conditions. If we write, by definition,

$$\lambda_i = \sum g_{ij}\epsilon_j,$$

the modul*

$$k_1\lambda_1 + k_2\lambda_2 + \cdots + k_n\lambda_n$$

is closed under multiplication on the left by every number σ of $\mathfrak{S}$. For

$$\sigma\kappa = \sum s_l k_i g_{ij} c_{ljr}\epsilon_r = \sum s_l k_i d_{lis}\lambda_s$$

is again in the modul.

If $\mathfrak{J}_1$ and $\mathfrak{J}_2$ are two left ideals of $\mathfrak{S}$ with bases $\lambda_1, \lambda_2, \cdots, \lambda_n$ and $\mu_1, \mu_2, \cdots, \mu_n$, respectively, the set of numbers

* Transactions of this Society, vol. 31 (1929), pp. 71–90.

$$\sum d_{ij}\lambda_i\mu_j,$$

where the d's range over $\mathfrak{P}$, is a linear form modul $\mathfrak{L}$. Since $\mathfrak{J}_1$ is closed under multiplication on the left by the numbers of $\mathfrak{S}$, the same is true of $\mathfrak{L}$. Hence $\mathfrak{L}$ is an ideal. Since it generalizes the concept of ideal product in algebraic number theory, it is called the *product* of the ideals $\mathfrak{J}_1$ and $\mathfrak{J}_2$ in that order.*

If $\mathfrak{J}_1\mathfrak{J}_2 = \mathfrak{K}$, the latter with basis $\nu_1, \nu_2, \cdots, \nu_n$, set

$$\lambda_i = \sum l_{ij}\epsilon_j, \quad \mu_i = \sum m_{ij}\epsilon_j, \quad \nu_i = \sum p_{ij}\epsilon_j.$$

Then there exist numbers h_{ijk} of $\mathfrak{P}$ such that

$$\lambda_i\mu_j = \sum h_{ijk}\nu_k.$$

That is,

$$\sum_{k,r,t} l_{ik}m_{jr}c_{krt}\epsilon_t = \sum_{k,t} h_{ijk}p_{kt}\epsilon_t,$$

or

$$\sum_{k,r} l_{ik}m_{jr}c_{krt} = \sum_{k} h_{ijk}p_{kt},$$

which can be written in either of the forms

$$G_1S(\mu_j) = H_{1j}G_3, \quad H_{1j} = (h_{rjs}),$$
$$G_2R^{\mathsf{T}}(\lambda_i) = H_{2i}G_3, \quad H_{2i} = (h_{irs}),$$

where G_1 is associated with $\mathfrak{J}_1$, G_2 with $\mathfrak{J}_2$, and G_3 with their product $\mathfrak{K}$. Similarly it may be shown that

$$G_3 = \sum_h K_{1h}G_1S(\mu_h) = \sum_j K_{2j}G_2R^{\mathsf{T}}(\lambda_j),$$

where K_{1h} and K_{2j} are matrices with elements in $\mathfrak{P}$. Thus G_3 may be readily determined as the g. c. r. d. of $G_1S(\mu_1)$, $G_1S(\mu_2)$, $\cdots$, $G_1S(\mu_n)$ or (if preferred) as the g. c. r. d. of $G_2R^{\mathsf{T}}(\lambda_1)$, $G_2R^{\mathsf{T}}(\lambda_2)$, $\cdots$, $G_2R^{\mathsf{T}}(\lambda_n)$.†

11. *Determination of a Minimal Basis.* If $\alpha_1, \alpha_2, \cdots, \alpha_k$ are numbers of $\mathfrak{S}$, the set of numbers

---

* G. Shover and C. C. MacDuffee, this Bulletin, vol. 37 (1931), pp.434–438.

† The problem of finding the minimal basis of the product of two quadratic ideals whose minimal bases are given was discussed by W. B. Carver, American Mathematical Monthly, vol. 18 (1911), pp. 81–87, with no simple results.

$$\xi_1\alpha_1 + \xi_2\alpha_2 + \cdots + \xi_k\alpha_k,$$

where the ξ's range over $\mathfrak{S}$, satisfies the definition of an ideal $\mathfrak{J}$. As stated before, this ideal is known to have a minimal basis $\lambda_1, \lambda_2, \cdots, \lambda_n$, such that the form

$$h_1\lambda_1 + h_2\lambda_2 + \cdots + h_n\lambda_n,$$

where the h's range over $\mathfrak{S}$, gives all the numbers of $\mathfrak{J}$. The proof of the existence of the minimal basis is existential, hinging upon the ability to select from an infinite set of numbers of $\mathfrak{S}$ one whose norm is minimal.

The effective determination of the minimal basis has been considered a problem of considerable difficulty. In 1917 M. Cipolla* gave a complete solution of this problem for quadratic fields, but even in this simple case both method and results are complicated.

The method of integral matrices affords a simple and easily stated solution of this problem in its general form.† Let $S(\alpha_i)$ be the second matrix of α_i, and let B be a matrix associated with the minimal basis $\lambda_1, \lambda_2, \cdots, \lambda_n$ to be found. The matrix B is a g. c. r. d. of $S(\alpha_1), S(\alpha_2), \cdots, S(\alpha_k)$.

Since $\mathfrak{J}$ is an ideal, it contains

$$\epsilon_r\alpha_i = \sum c_{rij}\beta_j, \quad (r, j = 1, \cdots, n;\ i = 1, \cdots, k).$$

That is, if $\alpha_i = \Sigma a_{ij}\epsilon_j$, $\beta_i = \Sigma b_{ij}\epsilon_j$,

$$\sum_{h,s} a_{ih}c_{rhs}\epsilon_s = \sum_{j,s} q_{rij}b_{js}\epsilon_s,$$

$$\sum_h a_{ih}c_{rhs} = \sum_j q_{rij}b_{js}, \quad (r, s, j = 1, \cdots, n;\ i = 1, \cdots, k).$$

In matric notation this is

$$S(\alpha_i) = Q_iB, \quad Q_i = (q_{ris}), \quad (i = 1, \cdots, k).$$

Hence B is a common right divisor of the matrices $S(\alpha_i)$.

Since every β_i is in $\mathfrak{J}$, there exist numbers

$$\pi_{rj} = \sum_h p_{rjh}\epsilon_h, \quad (r, h = 1, \cdots, n;\ j = 1, \cdots, k),$$

of $\mathfrak{S}$ such that

* Atti Accademia Catania, (5), vol. 10 (1917), No. 20.

† Mathematische Annalen, vol. 105 (1931), pp. 663–665.

$$\beta_r = \sum_j \pi_{rj}\alpha_j.$$

That is,

$$\sum_s b_{rs}\epsilon_s = \sum_{j,h,l,s} p_{rjh}a_{jl}c_{hls}\epsilon_s,$$

whence

$$b_{rs} = \sum_{j,h,l} p_{rjh}a_{jl}c_{hls}, \quad (r, s, h, l = 1, \cdots, n; j = 1, \cdots, k).$$

In matric notation this is

$$B = \sum P_j S(\alpha_j), \qquad P_j = (p_{rjs}).$$

Thus B is a g. c. r. d. of the matrices $S(\alpha_i)$.

It thus appears that a problem which is of great complexity when approached by the methods of elementary number theory and congruences can be given a very simple and natural treatment when the greatest common divisor theory of integral matrices is applied. Furthermore, in the case of algebraic fields at least, a constructive proof replaces an existential proof of the existence of a minimal basis of an ideal.

THE OHIO STATE UNIVERSITY

A NOTE ON THE DICKSON THEOREM ON UNIVERSAL TERNARIES*

BY A. A. ALBERT

1. *Introduction.* A form f with integer coefficients in integer variables is called *universal* if it represents *all* positive and negative integers. Evidently, since f is homogeneous, it represents zero for the variables all zero. In case $f=0$ for integral values of the variables not all zero f is called a zero form.

L. E. Dickson† has given a number-theoretic proof of his theorem that *every universal ternary quadratic form is a zero form.* But his proof is highly technical and consequently quite long and complicated. In the present note I shall give an almost trivial rational proof of Dickson's result. I shall also prove a generalization of his theorem for ternaries over any non-modular field F.

2. *Quadratic Forms over F.* Let F be any non-modular field and let $f(x_1, \cdots, x_n)$ be an n-ary quadratic form over F. Then we shall call f a *zero form* if $f=0$ for $x_1, \cdots, x_n$ in F and not all zero. We shall also say that, if every ρ in F is represented by f for $x_1, \cdots, x_n$ in F, the form f is universal over F.

It is well known‡ that there exists a non-singular linear transformation $x_i=\sum a_{ij}X_j$ with a_{ij} in F such that

$$f(x_1, \cdots, x_n) \equiv \phi(X_1, X_2, \cdots, X_n) \equiv \sum_{i=1}^{r} g_i X_i^2 + 0 \cdot \sum_{j=r+1}^{n} X_j^2,$$

with $g_i \neq 0$ in F. The integer r is the rank of f. Evidently f is a zero form if and only if ϕ is a zero form. But if $r<n$, the form ϕ vanishes for any X_n in F, if $X_1= \cdots =X_r=0$.

THEOREM 1. *Every n-ary of rank $r<n$ is a zero form. Every n-ary of rank n is equivalent to*

$$g_1X_1^2 + g_2X_2^2 + \cdots + g_nX_n^2, \qquad (g_i \text{ in } F),$$

with g_i all not zero.

---

* Presented to the Society, April 15, 1933.

† See his *Studies in the Theory of Numbers*, pp. 17–21.

‡ See Dickson, *Modern Algebraic Theories*, p. 69

3. *Proof of the Dickson Theorem.* Let $f(x, y, z)$ be a universal ternary. By Theorem 1 either f is a zero form of rank less than three or

$$(1) \qquad f(x, y, z) \equiv \phi(X, Y, Z) \equiv \alpha X^2 + \beta Y^2 - \gamma Z^2,$$

where α, β, γ are rational, $\alpha\beta\gamma \neq 0$, and X, Y, Z are linearly independent rational linear functions of x, y, z. Define

$$(2) \quad \delta \equiv \gamma(\alpha\beta)^{-1}, \quad a \equiv \alpha\delta, \quad b \equiv \beta\delta, \quad -ab = -(\alpha\beta\delta)\delta = -\gamma\delta,$$

so that, for a rational number $\delta \neq 0$,

$$(3) \qquad \delta f \equiv \delta\phi \equiv \psi(X, Y, Z) \equiv aX^2 + bY^2 - abZ^2.$$

Write $\delta = \delta_1\delta_2^{-1}$, where δ_1 and δ_2 are integers. Since f is universal, $f(x, y, z) = \delta_1\delta_2$ for integer x, y, z. Then if $x_0 = x\delta_1^{-1}$, $y_0 = y\delta_1^{-1}$, $z_0 = z\delta_1^{-1}$, we have $f(x_0, y_0, z_0) = \delta_1^{-2}\delta_1\delta_2 = \delta_2\delta_1^{-1} = \delta^{-1}$, for rational x_0, y_0, z_0. Hence we have proved the following fact.

LEMMA 1. *If f is universal, $\phi = \delta^{-1}$ for rational X, Y, Z.*

Let then $\delta^{-1} = \phi$, $\psi = \delta\phi = \delta\delta^{-1} = 1 = aX^2 + bY^2 - abZ^2$, and write as a consequence

$$(4) \qquad \xi \equiv 1 - aX^2 = bY^2 - abZ^2.$$

If $\xi = 0$, put $\eta = 1$, $\zeta = X$, so that

$$\psi(\xi, \eta, \zeta) = b \cdot 1^2 - ab \cdot X^2 = b(1 - aX^2) = b\xi = 0$$

for $\eta \neq 0$, and $\phi = \delta^{-1}\psi$ is a zero form. Hence f is a zero form, since $f = 0$ for rational x, y, z not all zero if and only if $f = 0$ for integers x, y, z, not all zero, since f is homogeneous.

Let then $\xi \neq 0$, and put $\eta = a(Z - XY)$, $\zeta = Y - aXZ$, so that

$$\begin{aligned} b\eta^2 - ab\zeta^2 &= b\left[a^2(Z^2 - 2XYZ + X^2Y^2)\right. \\ &\qquad \left. - a(Y^2 - 2aXYZ + a^2X^2Z^2)\right] \\ &= -ab\left[Y^2(1 - aX^2) - aZ^2(1 - aX^2)\right] \\ &= -a(1 - aX^2)(bY^2 - abZ^2) = -a\xi^2, \\ \delta\phi(\xi, \eta, \zeta) &\equiv a\xi^2 + b\eta^2 - ab\zeta^2 = 0, \end{aligned}$$

and $\phi(\xi, \eta, \zeta) = 0$ for $\xi \neq 0$. Hence again ϕ, and therefore also f, are zero forms, and we have proved the Dickson Theorem. The above proof is a rational proof holding for any field F so we have immediately the following result.

LEMMA 2. *If a ternary $f(x, y, z)$ with coefficients in F represents the associated quantity δ^{-1}, then f is a zero form.*

4. *Universal Ternaries over F.* We shall now prove the following theorem.

THEOREM 2. *A non-singular ternary quadratic form over F is universal over F if and only if it is a zero form.*

For let f be a zero form, so that $f(x, y, z) = 0$ for x, y, z not all zero and in F. Then

$$\delta\phi \equiv \psi(\xi, \eta, \zeta) \equiv a\xi^2 + b\eta^2 - ab\zeta^2 = 0$$

for ξ, η, ζ not all zero and in F. Let ρ be any quantity of F, $\sigma = \rho\delta$. If $\xi = 0$, then $b(\eta^2 - a\zeta^2) = 0$, whence $\eta^2 = a\zeta^2$, so that $\zeta\eta \neq 0$. Thus write $\xi_0 = \zeta\eta^{-1}$, from which $a\xi_0^2 = 1$. Put

$$X = 0, \qquad Y = \frac{\sigma + b^{-1}}{2}, \qquad Z = \frac{\sigma - b^{-1}}{2}\xi_0,$$

so that, since $1 = a\xi_0^2$,

$$\begin{aligned} 4\psi(X, Y, Z) &= b\left[(\sigma + b^{-1})^2 - (\sigma - b^{-1})^2 a\xi_0^2\right] \\ &= b\left[(\sigma + b^{-1})^2 - (\sigma - b^{-1})^2\right] \\ &= 4bb^{-1}\sigma = 4\sigma, \qquad \text{and} \qquad \psi = \sigma. \end{aligned}$$

Then $\phi = \delta^{-1}\sigma = \rho$ and hence $f = \rho$ for corresponding x, y, z in F.

Next let $\xi \neq 0$. Then $a + b(\eta\xi^{-1})^2 - ab(\zeta\xi^{-1})^2 = 0$, and if we write $\eta\xi^{-1} = a\zeta_0$, $\zeta\xi^{-1} = \eta_0$, we have $a + a^2b\zeta_0^2 - ab\eta_0^2 = 0$, $1 = b\eta_0^2 - ab\zeta_0^2$. Then put

$$X = \frac{\sigma + a^{-1}}{2}, \quad Y = \frac{\sigma - a^{-1}}{2}a\zeta_0, \quad Z = \frac{\sigma - a^{-1}}{2}\eta_0,$$

whence

$$\begin{aligned} 4\psi(X, Y, Z) &= a(\sigma + a^{-1})^2 + (ba^2\zeta_0^2 - ab\eta_0^2)(\sigma - a^{-1})^2 \\ &= a\left[(\sigma + a^{-1})^2 - (\sigma - a^{-1})^2\right] = 4aa^{-1}\sigma = 4\sigma, \\ \psi &= \sigma, \quad \phi = \delta^{-1}\sigma = \rho. \end{aligned}$$

Hence in this case also $f = \rho$ as desired, so that f is universal.

Conversely let f be universal. Then f represents δ^{-1} and, by Lemma 2, is a zero form. This proves* Theorem 2.

It is well known† that the determinant of the form $\phi(X, Y, Z)$ equivalent to f is h^2d, where h is the determinant of the transformation. Hence $-\alpha\beta\gamma = h^2d$, so that

$$\delta = \gamma(\alpha\beta)^{-1} = (\alpha\beta\gamma)(\alpha\beta)^{-2} = -dh^2(\alpha\beta)^{-2} = -dk^2,$$

where k is in F. Then

$$-df = \delta k^{-2}\phi = k^{-2}\psi(X, Y, Z) = \psi(\xi, \eta, \zeta),$$

for $X = k\xi \cdot X = k\eta \cdot Z = k\zeta$. Hence if f represents the negative of its determinant, the form $\psi = -df = (-d)^2$ represents d^2, and hence unity, and hence f is a zero form by Lemma 2. We may therefore replace Lemma 2 by the following statement.

THEOREM 3. *If f is a ternary with non-zero determinant d, then $-df(x, y, z) \equiv \psi(X, Y, Z) \equiv aX^2 + bY^2 - abZ^2$ for a suitable transformation. Also f is a universal zero form if and only if f represents $-d$.*

In particular the above Theorem 2 holds for the case where $F = R$, the field of all rational numbers. If, however, a is any rational number, then $a = b^{-2}c$, where b and c are integers. Obviously, if $f = a$ for rational x, y, z, then $f = c$ for rational x, y, z. Hence we have proved a partial converse to Dickson's theorem.

THEOREM 4. *A non-singular ternary quadratic form with integer coefficients is a zero form if and only if it represents all integers for rational values of its variables.*

THE UNIVERSITY OF CHICAGO

---

* It is evident that Theorem 2 is true if it can be proved for forms of type $\psi(X, Y, Z) \equiv aX^2 + bY^2 - abZ^2$. If $(1, i, j, ij)$, $i^2 = a$, $j = b$, $ji = -ij$, is a generalized quaternion algebra over F, then for $ab \neq 0$, this algebra is either a division algebra or a total matric algebra. If $q = Xi + Yj + Zij$, then $q^2 = \psi(X, Y, Z)$. Hence, if ψ is a zero form, the algebra Q is not a division algebra and there exists a two-rowed matrix whose square is σ so that ψ represents σ. The converse of Theorem 2 is similarly proved. It is in fact this linear algebra theorem (which has long been known to me) which gave me an immediate proof of Theorem 2 as soon as I discovered the reduction given by (1)–(3).

† See Dickson, *Modern Algebraic Theories*, pp. 64–70.

A SELF-DUAL SEPTIMIC POSSESSING SEVEN OF EACH KIND OF THE SIMPLE SINGULARITIES, AND AUTOPOLAR BY SEVEN RECTANGULAR HYPERBOLAS AND A CIRCLE*

BY D. C. DUNCAN

The equation of the locus in homogeneous rectangular coordinates, x, y, z, is developed under the assumptions:

i. The seven cusps are distributed at equal intervals about the unit-circle, their cuspidal tangents concurrent at (0, 0, 1).

ii. The curve is invariant under rotations about the origin through angles $2\pi k/7$, $(k=1, 2, 3, \cdots, 6)$.

iii. Each cuspidal tangent is an axis of symmetry.

As an immediate consequence of ii and iii, one node and one point of inflexion must lie on each cuspidal tangent.

The most general equation of a locus of the seventh order is

$$Af_7(x, y) + Bzf_6(x, y) + Cz^2f_5(x, y) + Dz^3f_4(x, y) + Ez^4f_3(x, y) + Fz^5f_2(x, y) + Gz^6f_1(x, y) + Hz^7 = 0.$$

The various $f_i(x, y)$ are homogeneous polynomial functions of degree i in x, y, which must be individually absolutely invariant under the rotations about the origin through angles $2k\pi/7$. Now $f_i(x, y)=0$ represents i straight lines through the origin. But it is manifestly impossible for i straight lines to be invariant under rotations of $2\pi/7$ if $i<7$, unless they are the isotropic lines $(x^2+y^2)^h=0$. Therefore we may write $f_6(x, y)\equiv(x^2+y^2)^3$, $f_4(x, y)\equiv(x^2+y^2)^2$, $f_2(x, y)\equiv(x^2+y^2)$, $C=0$, $E=0$, $F=0$, since the equation is to be real and to represent a real locus. Accordingly in polar coordinates the locus is represented by

$$A\rho^7 \cos(7\theta + \alpha) + B\rho^6 + D\rho^4 + F\rho^2 + H = 0.$$

Placing a cusp at $\rho=1$, $\theta=0$, causes α to vanish.

When $\theta=0$, the roots of $A\rho^7+B\rho^6+D\rho^4+F\rho^2+H=0$ must include a triple root for $\rho=1$, the cusp, and a double root for $\rho=a$, the node. It is next shown that for just one essential set of real values for A, B, D, F, H, the equation can have such roots. Assume

* Presented to the Society, March 18, 1933.

$$A\rho^7 + B\rho^6 + D\rho^4 + F\rho^2 + H \equiv A(\rho - 1)^3(\rho - a)^2(\rho^2 - p\rho + q).$$

Since the coefficients of ρ^5, ρ^3, ρ are zero, one has necessarily

$$\begin{aligned} a^2 + 6a + 3 + p(2a + 3) \quad &+ q \quad &= 0, \\ 3a^2 + 2a \quad + p(3a^2 + 6a + 1) &+ q(a^2 + 6a + 3) &= 0, \\ + pa^2 \quad &+ q(3a^2 + 2a) \quad &= 0. \end{aligned}$$

The consistency of these equations requires the vanishing of the determinant, that is,

$$\begin{vmatrix} a^2 + 6a + 3 & 2a + 3 & 1 \\ 3a^2 + 2a & 3a^2 + 6a + 1 & a^2 + 6a + 3 \\ 0 & a^2 & 3a^2 + 2a \end{vmatrix} = 0.$$

Expanding and factoring, one obtains

$$8a(a + 1)^2(3a^3 + 12a^2 + 16a + 4) = 0.$$

The vanishing of a would imply, contrary to hypothesis, a multiple point at the origin; the vanishing of $(a+1)$ is also to be discarded as the node cannot also lie on the unit circle for this would imply a circle meeting a septimic in 7 cusps and 7 nodes or 28 points, which is impossible. Hence the node is to be sought in the cubic factor. Fortunately this is reducible and yields the single root, $-\{(2/3)\ (7/2)^{1/3}+(4/3)+(7/6)\ (2/7)^{1/3}\}$. Moreover

$$p = -\frac{(3a + 2)(a^2 + 6a + 3)}{6(a + 1)^2}, \qquad q = \frac{a(a^2 + 6a + 3)}{6(a + 1)^2},$$

whence it is found that p^2-4q is definitely negative, indicating no other real intersection of the septimic with the x-axis.

Accordingly, for one single essential set of real values of A, B, D, F, H, the equation $A\rho^7 \cos 7\theta+B\rho^6+D\rho^4+F\rho^2+H=0$ represents a locus having the singularities situated according to the initial hypotheses. The nodes are found to be crunodes. It is clearly impossible for such a locus to consist of curves of lower order; hence the locus must be non-degenerate. Its equation in homogeneous rectangular coordinates is

$$(5(7/2)^{1/3} + 10 + 14(2/7)^{1/3})(x^7 - 21x^5y^2 + 35x^3y^4 - 7xy^6)$$
$$+ 7((13/3)(7/2)^{1/3} + 71/12 + (28/3)(2/7)^{1/3})(x^2 + y^2)^3z$$
$$- 21((43/8)(7/2)^{1/3} + 8 + (49/4)(2/7)^{1/3})(x^2 + y^2)^2z^3$$
$$+ 7((67/4)(7/2)^{1/3} + 101/4 + (77/2)(2/7)^{1/3})(x^2 + y^2)z^5$$
$$- ((953/24)(7/2)^{1/3} + 361/6 + (1099/12)(2/7)^{1/3})z^7 = 0.$$

Now if a polarity interchanges the cusps and inflexional tangents, cuspidal tangents and points of inflexion, the polarity will convert the entire locus into itself. This fact is due to the unique existence of the locus with the cusps in the specified position; the polar reciprocal curve possessing the same cusps must be coincident with the original septimic. The polarities are now exhibited.

The point and line coordinates of the following singularities are immediately deducible.

$$\begin{array}{lc} & x \quad : \quad y \quad : z \\ \text{Cusps} & \cos(2k\pi/7) : \sin(2k\pi/7) : 1, \\ \text{Pts. inflex.} & \cos(2k\pi/7) : \sin(2k\pi/7) : 0, \\ & u \quad : \quad v \quad : w \\ \text{Flex. tgts.} & \cos(2k\pi/7) : \sin(2k\pi/7) : c, \\ \text{Cusp. tgts.} & \cos(2k\pi/7) : \sin(2k\pi/7) : 0, \end{array}$$

$$\left(c = -\frac{52(7/2)^{1/3} + 71 + 112(2/7)^{1/3}}{12[5(7/2)^{1/3} + 10 + 14(2/7)^{1/3}]}\right), \quad (k = 0, 1, 2, \cdots, 6).$$

It is observed that only the following eight polarities interchange these singular elements:

$$\left.\begin{array}{l} \rho u = x, \\ \rho v = y, \\ \rho w = cz, \end{array}\right\} \quad \begin{array}{l} \rho u = x\cos(2k\pi/7) - y\sin(2k\pi/7), \\ \rho v = -x\sin(2k\pi/7) - y\cos(2k\pi/7), \\ \rho w = cz, \qquad\qquad (k = 0, 1, 2, \cdots, 6). \end{array}$$

Therefore the septimic is invariant under eight polarities, the first by the real circle $x^2+y^2+cz^2=0$, the others by the seven rectangular hyperbolas which touch this circle and whose principal axes coincide with the cuspidal tangents respectively. On account of the symmetry of the locus it is also invariant under six

correlations which are the products of the polarities ($k=1, 2, 3, 4, 5, 6$) with reflexion about any cuspidal tangent.

Correlations (Fourteen)

$$u : v : w$$

$$x : y : cz$$

$$x\cos(2k\pi/7) - y\sin(2k\pi/7) : -x\sin(2k\pi/7) - y\cos(2k\pi/7) : cz, \quad (k = 0, 1, \cdots, 6).$$

Polarities.

$$x\cos(2k\pi/7) + y\sin(2k\pi/7) : -x\sin(2k\pi/7) + y\cos(2k\pi/7) : cz, \quad (k = 1, 2, \cdots, 6).$$

Collineations (Fourteen)

$$\rho x = x\cos(2k\pi/7) \mp y\sin(2k\pi/7),$$

$$\rho y = x\sin(2k\pi/7) \pm y\cos(2k\pi/7),$$

$$\rho z = z, \qquad (k = 0, 1, 2, \cdots, 6).$$

The equations of the double tangents are now immediately obtainable by polarizing the nodes. They are

$$\{(2/3)(7/2)^{1/3} + 4/3 + (7/6)(2/7)^{1/3}\}x\cos(2k\pi/7) + y\sin(2k\pi/7) + \frac{52(7/2)^{1/3} + 71 + 112(2/7)^{1/3}}{12\{5(7/2)^{1/3} + 10 + 14(2/7)^{1/3}\}}z = 0, \quad (k = 0, 1, 2, \cdots, 6).$$

The line equation of the septimic is immediately obtained by replacing x, y, z by $u, v, w/c$.

The form of the curve is very approximately realized by drawing secant lines through alternate points of the seven points equally distributed about the unit circle, and then removing the chords from the interior of the circle. It is unipartite and of deficiency 1.

THE UNIVERSITY OF CALIFORNIA

ON IRREDUCIBLE SYSTEMS OF ALGEBRAIC DIFFERENTIAL EQUATIONS

BY E. GOURIN

Given two irreducible algebraic manifolds, if the first is a proper sub-manifold of the second, the first is of lower dimensionality than the second.* We prove here an analogous result for systems of algebraic differential equations.

Terminology and notation will be as in Ritt's monograph, *Differential Equations from the Algebraic Standpoint.*† Let Σ_1 be a non-trivial closed irreducible system. Let the unknowns be $u_1, \cdots, u_q, y_1, \cdots, y_p$, with the u_i arbitrary unknowns. We prove the following theorem.

THEOREM. *If the manifold of the non-trivial closed irreducible system Σ_1 is a proper sub-manifold of the manifold of another such system Σ_2, then either Σ_2 has a set of arbitrary unknowns of which $u_1, \cdots, u_q$ form a proper subset, or $u_1, \cdots, u_q$ form a complete set of arbitrary unknowns of the system Σ_2 and, for these arbitrary unknowns, resolvents of Σ_1 are of lower order than those of Σ_2.*‡

Because Σ_2 holds Σ_1, and Σ_1 is closed, every form of Σ_2 is contained in Σ_1. Then, certainly, Σ_2 cannot contain a form involving the u_i alone. Otherwise Σ_1 would contain such a form. Consequently, there exists in Σ_2 a set of arbitrary unknowns of which $u_1, \cdots, u_q$ form a subset. This subset is either a proper subset or a full set of arbitrary unknowns of Σ_2.

Let us assume, then, that this latter condition is satisfied, that is, $u_1, \cdots, u_q$ form a complete set of arbitrary unknowns of both systems Σ_1 and Σ_2.

In order to construct a resolvent of Σ_2 we choose two forms G and Q, with G not in Σ_2 and free of the y_i, such that, for any two distinct solutions of Σ_2 with the same u_i, such that G does not vanish for these solutions, Q yields two distinct functions

---

* Van der Waerden, *Moderne Algebra*, vol. 2, p. 63.

† Published by this Society, 1932.

‡ We assume that the associated field contains a non-constant function. This involves no loss of generality.

of x. Having selected two such forms, we adjoin to Σ_2 the form $w-Q$, where w is a new unknown, and obtain a system, say Λ_2. It is clear, however, that the above described properties of the forms G and Q with respect to the system Σ_2 will remain undisturbed when tested with respect to the system Σ_1. Hence, the same forms G and Q may be utilized in the construction of the resolvent of Σ_1. Accordingly, we adjoin the same form $w-Q$ to Σ_1, obtaining a system Λ_1.

Let Ω_1 and Ω_2 be the two systems of forms in w, the u_i, and y_i, which vanish for all solutions of Λ_1 and Λ_2, respectively. The two systems Ω_1 and Ω_2 are closed and irreducible. Furthermore, any form of Ω_2 is contained in Ω_1.

We now list the unknowns in Ω_1 in the order $u_1, \cdots, u_q; w; y_1, \cdots, y_p$, and take a basic set for Ω_1

$$A, A_1, \cdots, A_p, \tag{1}$$

in which $w, y_1, \cdots, y_p$ are introduced in succession and in which A is algebraically irreducible. The equation $A=0$ is a resolvent of Ω_1. Each A_k is a form in w, the u_i, and y_k alone, is of order 0 in y_k, and of the first degree in y_k.

Similarly we proceed with Ω_2 and take a basic set

$$B, B_1, \cdots, B_p, \tag{2}$$

with B algebraically irreducible. Then $B=0$ is a resolvent of Ω_2 and each B_k is of the same structure as the corresponding A_k.

It is clear that A cannot be of higher order in w than B. Let us assume, then, that A and B are of the same order in w. Let I be the initial of A. Then, for an appropriate non-negative integer t, we have the identity

$$I^tB = AC + D, \tag{3}$$

where C and D are forms in w and the u_i, D being of lower rank in w than A. Since D vanishes for every solution of Ω_1, D is contained in Ω_1. Hence $D \equiv 0$.

It follows, therefore, from (3) that I^tB is divisible by A. Because A is algebraically irreducible, and I and A are relatively prime, B must be divisible by A. Let $B=A\cdot K$. Then K is a function of x, the form B being algebraically irreducible. This implies that in (1) we may replace A by B, because, in choosing this basic set, we may select for A among the forms of Ω_1 in-

volving w and the u_i alone, any algebraically irreducible form which is of the least rank in w. Such a form, however, was found to be a member of Ω_2 and hence may be identified with the form B.

B_1 is contained in Ω_1. It is reduced with respect to B and, of all such forms in Ω_1 in the u_i, w, and y_1, it certainly has a lowest rank. Consequently we may replace A_1, in (1), by B_1. Continuing, we find that (2) is a basic set for Ω_1. Then Σ_1 and Σ_2 are identical. This contradiction proves that A is of lower order in w than B and establishes our theorem.

COLUMBIA UNIVERSITY

AXIOM C OF HAUSDORFF AND THE PROPERTY OF BOREL-LEBESGUE*

BY SELBY ROBINSON†

1. *Introduction.* This is a study in an abstract space (P, K) of the Hausdorff‡ property C which may be expressed in the form *the interior of every set is an open set.* A point p of the space P is interior to a set V, if p is a point of V and is not a K-point (point of accumulation, limit point) of any subset of $C(V)$. An open set is one all of whose points are interior points. We say that space (P, K) has property B of Hausdorff if and only if any point p which is interior to each of two sets is interior to their logical product; we shall designate as the open set B property, the weaker property: the product of two open sets is an open set.§ By the Hausdorff property D we shall mean that any two points are respectively interior to sets which are disjoined, while in the open set D property the points are required to be in disjoined open sets. The Borel and Borel-Lebesgue properties take three non-equivalent forms in spaces not having property C. These three forms coincide if property C is present as do the two forms of property B and of property D. In §3 we consider three

---

* Presented to the Society, October 29, 1932.

† National Research Fellow.

‡ F. Hausdorff, *Grundzüge der Mengenlehre*, first edition, 1914, p. 213.

§ Chittenden chose the open set B property as the one to designate as the Hausdorff B property. See Transactions of this Society, vol. 31 (1929), p. 315.

set functions introduced by Chittenden (loc. cit.) which are analogous, respectively, to the three forms of the Borel and Borel-Lebesgue properties. We show that a necessary and sufficient condition for property C is that some two of these three functions coincide for all sets of the space. In §4 we prove that a space having properties B and D and a local form of the Borel-Lebesgue property also has property C.

2. *The Borel Property.* A set E has the Borel property if every enumerable family of sets F which properly covers E (that is, every point of E is interior to some set of the family F) has a finite sub-family which also properly covers E.* A weaker property asserts that any enumerable proper covering F of E has a finite sub-family F_1 whose sets contain all points of E (F_1 covers E simply). By the still weaker open set Borel property, an enumerable family of open sets which covers E is reducible to a finite one. In a euclidean space the sets having these properties are those which are closed and bounded; while the bounded sets are the sets E having the property that every enumerable family of open sets which covers the space P has a finite sub-family which covers E. The latter property implies that every infinite sub-set Q of E has a V-point† in the space. If the space has property B and no point has another as a limit point, a point which is a V-point of an enumerably infinite set is also a nuclear point of that set. But I have shown that a necessary and sufficient condition that every enumerably infinite subset of E have a nuclear point is that every enumerable proper covering of P be reducible to a finite simple covering‡ of E. Thus the latter property is equivalent to the reducibility of enumerable families of open sets which cover P to finite coverings of E, in case property B holds and no point has another as a limit point. For the case $E=P$, we formulate this theorem as follows.

THEOREM 1. *A necessary and sufficient condition that a space (P, K) in which property B holds and no point has another as*

* The Borel-Lebesgue property states that proper coverings of any power are reducible to finite ones.

† A point p is a V-point of a set Q if for every set W to which p is interior $Q \cdot W$ contains a point other than p; a nuclear point if the cardinal number of $Q \cdot W$ is the same as that of Q.

‡ *Covering theorems in general topology*, American Journal of Mathematics, vol. 55 (1933), Theorem 2.

limit point, have the open set Borel property is that every enumerable proper covering be reducible to a finite simple one.

In particular, the properties are equivalent in a Fréchet $\mathcal{L}$-space. Chittenden has given an example* of a Fréchet $\mathcal{L}$-space having the properties of Theorem 1 but not the Borel property, and we give another in §4. The following example of a space having the open set Borel property but not the "proper to simple" form, shows that the condition, no single point has a limit point, is essential in Theorem 1. The space consists of a sequence of points $\{p_n\}$ and the sets having p_n as a limit point are exactly those which contain p_{n-1}.

The following is an example of a set E of a Fréchet $\mathcal{L}$-space which has the open set Borel property but not the "proper to simple" form. The space consists of two concentric circles D and E. The sets which have any point p for a limit point are exactly those which ordinarily would have as a limit point the corresponding point q of the other circle. Any enumerable family of sets which are open sets in the space $P=D+E$ has a finite sub-family which covers E; but if E is considered as a space in itself, no set has a limit point and the open set Borel theorem is not valid. In contrast to this, it is a stronger condition for a subset E to have the Borel property than for E considered as a space to have the Borel property. For consider a Fréchet $\mathcal{L}$-space P consisting of a plane with the following limit relations. Any point p whose x-coordinate is irrational is a limit point of those sets whose intersection with the horizontal line containing p would ordinarily have p as a limit point; while a point p whose x-coordinate is rational, is a limit point of the sets whose intersection either with the horizontal or the vertical line containing p would ordinarily have p for a limit point. A closed segment E of a horizontal line does not have the Borel property, but does have it if E is considered as a space. Finally in the "proper to simple" case, it makes no difference whether E be considered as a space or as a subset.

3. *Set Functions.* Analogous to $V(E)$ is the set $O(E)$ of all points p such that every open set containing p contains a point of E other than p, and the set $H(E)$ of all points p such that

* This Bulletin, vol. 25 (1918), p. 64.

every set to which p is interior has in its interior a point of E other than p. If we omit the phrase "other than p" from the definitions of $H(E)$, $V(E)$, and $O(E)$, we get, respectively, the sets $H_1(E)=H(E)+E$, $M(E)=V(E)+E=L(E)+E$, and $O_1(E)=O(E)+E$. Evidently $H(E)\leqq V(E)\leqq O(E)$ and $H_1(E)\leqq M(E)\leqq O_1(E)$. Since property C means* that $I\,I(E)\equiv I(E)$, and since $M(E)\equiv CIC(E)$, a necessary and sufficient condition for property† C is that $MM(E)\equiv M(E)$. But in every space‡

$$H_1H_1(E)\equiv H_1(E) \text{ and } O_1O_1(E)\equiv O_1(E).$$

Therefore if $M(E)=H_1(E)$ or $M(E)=O_1(E)$, property C holds.

THEOREM 2. *If property C holds, $H_1(E)=M(E)=O_1(E)$. Conversely, if any two of these three set functions coincide for all sets of the space, property C holds.*

From Theorem 2 we can deduce another necessary and sufficient condition for property C; namely, that some two of the set functions $H(E)$, $V(E)$, and $O(E)$ coincide for all sets of the space.

4. *The Borel-Lebesgue Property.* Property C holds if the space has the Borel property and every subset having that property is completely closed, or if the space has the Lindelöf property and every set having that property is completely closed; or if the space has the Borel-Lebesgue property and every set having that property is completely closed. If property B is valid, every point is interior to a set having the property of Borel-Lebesgue, and every set having that property is completely closed, then

---

* Where $I(E)$ is the set of all interior points of E, and the symbol $\equiv$ means that the formula holds for every set E of the space.

† This property is equivalent to the closure of derived sets in any space having the first three properties of Riesz, but the two properties are independent in V-spaces. If $M(Q)=Q$, we say that Q is completely closed or L-closed.

‡ These equalities are special cases of the relation $I_I(E)=I_I(E+I_I(E))>I_II_I(E)$, where I is now any set function and $I_I(E)$ is the set of points p such that $I(W)>p$ implies $E\cdot I(W)\neq 0$. See Robinson, loc. cit., §5. If $E\cdot I(W)$ contains at least μ points, we say that $p<I_I^{\mu}(E)$. Then $I_I^{\mu}(E)>I_II_I^{\mu}(E)$. From this we can prove that if property C holds, the set of points p such that every set to which p is interior contains μ points of E, is a completely closed set. For $\mu=\aleph_1$, this means that points of condensation of any set E form a completely closed set.

a proof similar to that of Theorem 3 would show that property C holds.

THEOREM 3. *If for each point p of a space (P, K) there is a set U_p to which p is interior and such that every proper covering of P is reducible to a finite proper covering of U_p, and if properties B and D hold, then property C also holds.*

PROOF. We suppose p is not an M-point of the set E and prove it is not an M-point of $M(E)$. For each point q of $M(E)$, let V_q and W_q be disjoined sets to which p and q are respectively interior. Then the W_q's together with $C(E)$ form a proper covering of P, which is reducible to a finite proper covering of U_p. Since $C(E)$ covers no points of $M(E)$, $U_p \cdot M(E)$ is properly covered by sets W_{q_i} which form a finite subfamily of the sets W_q. Therefore the product $U_{p_i} \cdot \Pi V_{q_i}$ to which p is interior, contains no points of $M(E)$.

Consider a space consisting of the circumference of a circle. The limit points of a set E are the points p such that there are points of $E-p$ either (1) arbitrarily near p in the direction clockwise from p, or (2) arbitrarily near q(the point opposite p) in the direction counter clockwise from q. This space does not have property C. It satisfies properties B and D but not the other hypothesis of Theorem 3. However, every proper covering is reducible to a finite simple covering.*

By a proof like that of Theorem 3 we can establish the following theorem.

THEOREM 4. *In a space (P, K) having properties B and D, any set E such that every proper covering of E is reducible to a finite simple one, is completely closed.*

Theorems 5 and 6 which follow are corollaries of Theorem 4, while Theorem 7 is a consequence of a similar theorem in which

* In this Bulletin, vol. 37 (1931), p. 630, I stated erroneously that the above space is a Hausdorff space having the Borel-Lebesgue property. By a modification we can secure the space which we desired, a separable Hausdorff space having the Borel-Lebesgue property and satisfying the first but not the second enumerability axiom. Let the space consist of two concentric circles. Let the limit points of a set E be as above, except that for any point p on either circle, the point q shall be the corresponding point on the other circle. In this example (and in one of §2), we could just as well have taken two parallel closed segments, except that we desired to make the space homogeneous.

the open set B and D properties are assumed and the set E is only required to have the open set Borel-Lebesgue property.

THEOREM 5. *In a space (P, K) having properties B and D, a necessary and sufficient condition that a set E have the property of Borel-Lebesgue is that E be L-closed and every infinite proper covering of P be reducible to a proper covering of E of lower power.*

THEOREM 6. *In a space P having properties B and D, a necessary and sufficient condition that every infinite proper covering of a set E be reducible to a finite simple one is that E be L-closed and every infinite proper covering of P be reducible to a finite simple covering of E.*

THEOREM 7. *In a space P having the open set form of properties B and D, a necessary and sufficient condition that every infinite family of open sets which covers E be reducible is that E be L-closed and every infinite family of open sets which covers P be reducible to a covering of E of lower power.*

Each of the reducibility properties of Theorems 3, 4, and 5 is equivalent to a nuclearity property and to a property relating to the closure of decreasing sequences of sets.* If these nuclearity and closure properties are substituted for the reducibility properties in Theorems 3, 4, and 5, additional theorems are secured. For example, in a space (P, K) having properties B and D, a necessary and sufficient condition that every infinite subset of the set E have a proper nuclear point in E is that E be L-closed and every infinite subset of E have a proper nuclear point in the space.

If the space satisfies the first enumerability axiom, the reducibility of enumerably infinite coverings can be substituted in Theorems 3–7 for the reducibility of coverings of any infinite power.

PRINCETON UNIVERSITY

* See Robinson, loc. cit.

AXIOMATIC DEFINITIONS OF PERFECTLY SEPARABLE METRIC SPACES*

BY GARRETT BIRKHOFF

1. *Introduction.* Among topological spaces, perfectly separable metric spaces (abbreviated PSM-spaces) play a major† role; therefore their axiomatic definition is a problem of peculiar interest. We shall discuss axiomatic definitions of PSM-spaces.

A *topological space* is‡ a class $\mathfrak{S}$ of "points" x, and an operation K assigning to any set S of points of $\mathfrak{S}$, a "derived" set $K(S)$. In practice, K is defined in one of three ways:

I. By ascribing to every point x of $\mathfrak{S}$ a family of "neighborhoods of x," denoted by V_x with variable superscript. Here $K(S) \supset x$, if and only if no $(V_x - x) \cdot S$ is vacuous.

II. By ascribing to every ordered pair (x, y) of points of $\mathfrak{S}$, a number $\overline{xy} \cdot \overline{xy} > 0$ if $x \neq y$; $\overline{xx} = 0$. Here $K(S) \supset x$, if and only if g.l.b. $_{S \supset x \neq y} \overline{xy} = 0$.

III. By ascribing to every point x of $\mathfrak{S}$ a family of infinite sequences $\{x_k\}$ of distinct points of $\mathfrak{S}$, which "converge to x" (abbreviated $x_k \to x$). Here $K(S) \supset x$, if and only if we can find a sequence in S converging to x.

Spaces defined as in I are known as spaces (V); we shall call spaces defined as in II spaces (Q), and those defined as in III, spaces (Λ).

A system of neighborhood axioms, such that any space (V) satisfying it is a PSM-space, has been deduced by Hausdorff-Urysohn-Tychonoff.§ With slight modifications, the axioms are satisfied by every representation of a PSM-space as a space (V). It is therefore futile to attempt to weaken them except as regards interdependence.

---

* Adapted from a Harvard Undergraduate Honors Thesis.

† For they are precisely the class used in function theory. See Mathematische Annalen, vol. 92 (1924), p. 302.

‡ Fréchet, *Les Espaces Abstraits*, 1928, p. 167.

§ See Hausdorff's *Grundzüge der Mengenlehre*, 1914, Chaps. 7–8; Urysohn, *Zum Metrisationsproblem*, Mathematische Annalen, vol. 94 (1925), p. 309; Tychonoff, *Über ein Satz*, etc., Mathematische Annalen, vol. 95 (1925), p. 139.

Systems of distance axioms, all spaces (Q) that satisfy which are metric spaces, have been deduced by Chittenden-Niemytski-Wilson.* We shall proceed by another line to find a system of distance axioms, effectively weaker under the additional hypothesis of perfect separability—satisfied in fact by every representation of a PSM-space as a space (Q). No system of convergence axioms, all spaces (Λ) that satisfy which are PSM-spaces, has, I believe, been set up hitherto. We shall construct such a system, showing in what sense it is necessary as well as sufficient.

2. *Spaces* (Λ) *and Spaces* (Q). The class of spaces (Λ) is quite evidently topologically equivalent† (abbreviated TE) to the class of spaces (V) in which, if $K(S) \supset x$, S contains an enumerable subset S' such that $K(S') \supset x$. Suppose that in addition we can so choose S' that $\{y_k\} \subset S'$ implies $K(\{y_k\}) \supset x$. We can make the class of such sets S' our fundamental class of convergent sequences. We then have a space (Λ) in which

(A) *If* $x_k \to x$, *and* $\{y_k\} \subset \{x_k\}$, *then* $y_k \to x$. *Conversely if, given* $\{y_k\} \subset \{x_k\}$, *we can so choose* $\{z_k\} \subset \{y_k\}$ *that* $z_k \to x$, *then* $x_k \to x$.

THEOREM 1. *Let* $(\mathfrak{S}, K)$ *be a space* (Λ) *satisfying*

(B) *If* $x_k \to x$, *and* $\{y_k\} \subset \{x_k\}$, *then we can so choose* $\{z_k\} \subset \{y_k\}$ *that* $z_k \to x$.

Then $(\mathfrak{S}, K)$ *has a* TE *representation as a space* (Λ) *satisfying* (A).

For add to the primitive families of convergent sequences of $(\mathfrak{S}, K)$, all sequences satisfying the second half of (A). Then K will be unchanged, yet (A) will be satisfied by the augmented families of convergent sequences. It is noteworthy that any further augmentation either contradicts (B) or changes K.

A space is called *locally separable* if and only if it is TE to a space (V) in which the family of neighborhoods ascribed to any x can be enumerated: $V_x^1, V_x^2, V_x^3, \cdots$. Every space (Q) is locally separable; let $V_x^n \supset y$ if and only if $\overline{xy} < 1/n$. Every space

* For latest results see Niemytski's *On the third axiom of metric space*, Transactions of this Society, vol. 29 (1927), p. 666; also W. A. Wilson's two articles in the American Journal, vol. 53 (1931).

† Two topological spaces are called "topologically equivalent" if and only if a (1, 1)-correspondence which preserves K can be established between their points.

(Λ) is* a space (V), but not necessarily a locally separable one.

Let $(\mathfrak{S}, K)$ be a locally separable space (Λ), and define a pseudo-distance $\tilde{xy}$ as follows: $\tilde{xx}=0$; and $\tilde{xy}=1/n$ $[x\neq y]$, if V_x^n is the first V_x in order of enumeration not† containing y. If g.l.b.$_{S\supset y\neq x}\tilde{xy}=0$, then plainly $K(S)\supset x$; i.e., given n, $y\subset S$ exists contained in at least the first n neighborhood V_x^k. Suppose $K(S)\supset x$, yet g.l.b.$_{S\supset y\neq x}\tilde{xy}>0$. Hence $S\supset x_k\to x$, yet g.l.b. $\tilde{xx}_k>0$; therefore that some V_x^n excludes an infinity of the x_k. Extracting the subsequence not in V_x^n, we obtain an exception to (A).

Conversely, in a space (Q), let us write $x_k\to x$ if and only if $\overline{xx}_k\to 0$. It follows from the theory of limits, that $K(S)\supset y$ if and only if an infinite sequence $\{x_k\}$ exists in S, satisfying $x_k\to y$. Moreover the pseudo-convergences just defined satisfy (A). We may therefore state the following theorem.

THEOREM 2. *The class of locally separable spaces* (Λ) *satisfying* (A) *is* TE *to the class of spaces* (Q). *Moreover* $\overline{xx}_k\to 0$ *is equivalent to* $x_k\to x$.

3. *Locally Separable Hausdorff Spaces.* Consider $(\mathfrak{S}, K)$, any space (Q) satisfying the two conditions

(α) If $\overline{xz}_i+\overline{yz}_i\to 0$, then $x=y$.

(β) If $\overline{xx}_i\to 0$, and $\overline{x_ix_j}^i\to 0$ identically in i, then $N(i)$ exists so large that $j(i)>N(i)$ implies $\overline{xx}_{j(i)}\to 0$.

LEMMA. $K\{K(S)\}\subset K(S)$ *for any set* S *in* $(\mathfrak{S}, K)$.

For suppose $x\subset K\{K(S)\}$; we can evidently so choose $x_i\subset K(S)$ that $\overline{xx}_i\to 0$. But since $x_i\subset K(S)$, we can so choose $x_j^i\neq x$ in S that $\overline{x_ix_j}^i\to 0$, irrespective of i. Hence by (β), g.l.b.$_{S\supset x_j^i\neq x}\overline{xx_j}^i=0$, and $x\subset K(S)$.

Let us define V_x^n $[x\subset\mathfrak{S};\ n=1, 2, 3, \cdots]$ as the set of all points y which satisfy $\overline{xy}>1/n$; let $W_x^n=V_x^n-K(\mathfrak{S}-V_x^n)$. Then $V_x^n\supset x$, but x is not $\subset K(\mathfrak{S}-V_x^n)$, so that the following statement will be true.

(a) To every x corresponds a W_x^n, and every W_x^n contains x. Moreover the theory of limits shows us that in any space (Q), $K(S+T)=K(S)+K(T)$. It follows by this and the Lemma that

* Since $S\supset S'$ implies $K(S)\supset K(S')$; see *Les Espaces Abstraits*, p. 179.

† If $x\neq y$, *some* $V_x^n\supset y$; otherwise $K(y)\supset x$, and since y contains no infinite sequence, this is impossible.

$$(\phi) \qquad K(\mathfrak{S} - W_x^i) = K(\mathfrak{S} - V_x^n) + K\{K(\mathfrak{S} - V_x^n)\}$$
$$= K(\mathfrak{S} - V_x^n) \subset \mathfrak{S} - W_x^n.$$

Now suppose that some $(W_x^n - x) \cdot S$ is vacuous, so that $(\mathfrak{S} - W_x^n) + x \supset S$. It follows that $K(\mathfrak{S} - W_x^n) + K(x) \supset K(S)$; hence, using (ϕ) and the fact that $K(x)$ is null, we see that $(\mathfrak{S} - W_x^n) \supset K(S)$, and, by (a), $K(S)$ is not $\supset x$. Conversely, if no $(W_x^n - x) \cdot S$ is vacuous, certainly no $(V_x^n - x) \cdot S$ is vacuous, and $K(S) \supset x$. That is, the W_x^n are neighborhoods of x in the sense of I.

(b) Let W_x^i and W_x^j be given. Then $W_x^{i+j} \subset W_x^i \cdot W_x^j$ exists.

(c) If $y \subset W_x^n$, then y is not $\subset (\mathfrak{S} - W_x^n)$, whence by (ϕ), we have y is not $\subset K(\mathfrak{S} - W_x^n)$. Therefore W_y^n exists, making $W_y^n \cdot (\mathfrak{S} - W_x^n)$ null, and hence contained in W_x^n.

(d) Let $x \neq y$ be given. Then W_x^n and W_y^n exist without common point; otherwise $z_n \subset W_x^n$, W_y^n exists for every n, and $\overline{xz}_n + \overline{yz}_n \to 0$, contradicting (α).

Conversely, setting up a pseudo-distance just as in the case of locally separable spaces (Λ), we obtain a TE space (Q) from any space (V) which is locally separable and satisfies (a) and (b). Further, we can obtain (α) from (d) and (β) from (c). Therefore we may state the following result.

THEOREM 3. *The class of spaces* (Q) *satisfying* (α) *and* (β) *is* TE *to the class of locally separable Hausdorff spaces.*

4. *Two Further Conditions.* A topological space is said to be *regular** if and only if to every point x and every closed set S not containing x correspond disjoint open sets $U \supset x$ and $V \supset S$. Suppose $(\mathfrak{S}, K)$ is a space (Q) satisfying (α) and (β) but not regular, and let x and S defined as above be contained in no two disjoint open sets. Let finally W be defined as in §3.

Every W_x^i must have either a point or a limit point lying in S; otherwise we could enclose each point of S in an open set disjoint to W_x^i, and the sum of these open sets would be an open set containing S and disjoint to $W_x^i \supset x$. Since $K(S)$ is not $\supset x$, $W_x^i \cdot S$ is vacuous for $i > M(S, x)$, and to $i > M(S, x)$ correspond $x_j^i \subset W_x^i$ such that $x_j^i \to x_i \subset S$. But since x is not $\supset S \supset K(S)$, g.l.b.$_{S \supset x_i} \overline{xx}_j^i > 0$. That is, if we set $N(i) = 0$, there exists an exception to this statement:

* E. W. Chittenden, this Bulletin, vol. 33 (1927), p. 20. The term is due to Alexandroff and Urysohn, Mathematische Annalen, vol. 92 (1924), p. 263.

(β') If $N(i)$ exists so large that $j(i)>N(i)$ implies $\overline{xx}^i_{j(i)}\to 0$, and $\overline{x_ix}^i_j\to 0$ identically in i, then $\overline{x\bar{x}}_i\to 0$.

We have therefore proved the following fact.

THEOREM 4. *Every space* (Q) *satisfying* (α), (β), *and* (β') *is* TE *to a regular locally separable Hausdorff space.**

We shall call a space *perfectly separable* if and only if we have:

(γ) There exist open sets $N_1, N_2, N_3, \cdots$, of which any open set is the sum of a subclass.

THEOREM 5. *If a Hausdorff space satisfies* (γ), *it is perfectly separable in the sense of Hausdorff, that is, we can take the* N_i *for neighborhoods; and conversely.*

Let N_i be a neighborhood of x if and only if $N_i \supset x$, and observe (1) that since each N_i is open, it contains at least one neighborhood of every point contained in itself; (2) that since by (c) every V_x is open, $V_x=\sum N_i$ summed over a suitable range of i,—whence since in addition by (A) $V_x \supset x$, it follows that $V_x \supset N_i \supset x$ is satisfied by at least one N_i. It results from (1) and (2) that the N_i define K correctly, while (A) – (D) can be easily deduced from (1), (2), and the fact that the N_i are open sets containing the points of which they are neighborhoods. To prove the converse, merely set the V_x as N_i.

THEOREM 6. *If the derived sets of a space* (Λ) *are closed, and it satisfies* (γ), *it is locally separable.*

For in this case the interior points of neighborhoods of a point form an equivalent† system of neighborhoods and are open; from this point we can proceed as in the proof of the first part of Theorem 5.

If we note that $\overline{x\bar{x}}_i\to 0$ means simply that $k_i<k_{i+1}$ implies $K(\sum_1^\infty x_{k_i}) \supset x$, we see immediately that conditions (α)–(γ) are topologically invariant. Likewise, since under any metric‡ representation $\overline{xz}_i+\overline{yz}_i \geqq \overline{xy}$, while we can so choose $N(i)$ that $\overline{x_ix}^i_j<1/i$, whence $|\overline{xx}_i-\overline{xx}^i_j|<1/i$, for $j>N(i)$, we see that (α)–(γ) hold under the metric representation of a PSM-space.

* The converse is also true.

† *Les Espaces Abstraits*, p. 173.

‡ A metric representation is a representation as a space (Q) in which a. $\overline{xy}=\overline{yx}$, b. $\overline{xy}+\overline{yz}\geqq\overline{xz}$.

Hence (α)-(γ) hold under any representation of a PSM-space as a space (Q). However, Tychonoff† has proved that a necessary and sufficient condition that a perfectly separable Hausdorff space be metric is that it be regular. Applying Theorems 4 and 5, we may therefore state the following conclusion.

THEOREM 7. *A necessary and sufficient condition that a space* (Q) *be a* PSM-*space is that* (α), (β), (β'), *and* (γ) *hold.*

5. *Conclusions.* Inasmuch as every space (Q) is locally separable, and $\aleph^2 = \aleph$, it might be conjectured that (γ) was replaceable by the following neater condition:

(γ') $\Gamma = x_1 + x_2 + x_3 + \cdots$ exists in $\mathfrak{S}$, and $K(\Gamma) = \mathfrak{S}$.

THEOREM 8. *There exists a space* (Q) *which is not a* PSM-*space, yet which satisfies* (α), (β), (β'), *and* (γ').

For let $\mathfrak{S}$ consist of all points $x[0<x<1]$, and let $\overline{xy} = (x-y)$ or 1, according as $x \geqq y$ or $x<y$. That $(\mathfrak{S}, K)$ satisfies the hypotheses is fairly obvious; the only difficulty is to show that it is not perfectly separable. To show this, remark that *every* V_{x_0} contains a set $x_0 > x > x_0 - \epsilon$, and some $V_{x_0}^*$ excludes the entire set $x < x_0$ as well. Now $V_x^* = V_y^*$ only if $x = y$. Hence the neighborhoods are certainly not enumerable.

Let us disprove in the same way a second conceivable conjecture. Let $\mathfrak{S}$ consist of two classes of points, points p_k and points $q_k[0=k=1]$. Let $\overline{q_iq_j} = \overline{q_ip_j} = \overline{p_kq_k} = \overline{q_kp_k} = 1$, and $\overline{p_ip_j} = \overline{p_iq_j} = |i-j|, [i \neq j]$. It is easy to verify, using Theorem 4, that the space is a compact regular locally separable Hausdorff space. It is not perfectly separable, as it contains more than countable isolated points. Therefore it is not metric; every compact metric space is‡ perfectly separable. Hence we have the following fact.

THEOREM 9. *There exists a compact regular locally separable Hausdorff space which is not metric.*

By simple ad hoc constructions we can show the independence of (α), $(\beta)+(\beta')$, and (γ) as conditions on a space (Q); the last

† Mathematische Annalen, vol. 95 (1925), p. 139.

‡ See Urysohn, *Zum Metrisationsproblem*, Mathematische Annalen, vol. 94 (1925), p. 309.

is incidentally proved by Theorem 8. From this and Theorem 7 we have the following result.

THEOREM 10. *The following distance axioms determine which spaces* (Q) *are* PSM-*spaces*:

(α) *If* $\overline{xz}_i+\overline{yz}_i\to 0$, *then* $x=y$.

(β'') *Let* $\overline{x_ix^j}\to 0$ *identically in* i. *Then* $\overline{xx}_i\to 0$ *if and only if* $N(i)$ *exists, so large that* $j(i)>N(i)$ *implies* $\overline{xx}^i_{j(i)}\to 0$.

(γ) *There exist open sets*† $N_1, N_2, N_3, \cdots$, *of which any open set is the sum of a subclass.*

Employing Theorem 2 to replace (α) and (β'') by equivalent axioms, and suppressing the hypothesis of local separability as being implied in the others (see Theorem 6), we obtain the following theorem.

THEOREM 11. *Taken together with* (γ), *the following convergence axioms determine which spaces* (Λ) *satisfying* (A) *are* PSM-*spaces*:

(B) *If* $z_i\to x$, *and* $z_i\to y$, *then* $x=y$.

(C) *Let* $x_i^j\to x_i$ *identically in* i. *Then* $x_i\to x$ *if and only if* $N(i)$ *exists, so large that* $j(i)>N(i)$ *implies* $x^i_{j(i)}\to x$.

Finally, let $(\mathfrak{S}, K)$ be any *topological space*. Let us make the *definition*, $\{x_k\}\to x$ if and only if $\{y_k\}\subset\{x_k\}$—all points of the same infinite sequence being assumed distinct—implies $K(\{y_k\})\supset x$. Then, since (A') and the new definition of convergence imply (A), we may state the following result.

THEOREM 12. *The following convergence axioms determine which topological spaces are* PSM-*spaces*:

(A') $K(S)\supset x$ *if and only if* $\{x_k\}\subset S$ *exists, such that* $\{x_k\}\to x$.

(B) *If* $z_i\to x$, *and* $z_i\to y$, *then* $x=y$.

(C) *Let* $x_i^j\to x_i$ *identically in* i. *Then* $x_i\to x$ *if and only if* $N(i)$ *exists, so large that* $j(i)>N(i)$ *implies* $x^i_{j(i)}\to x$.

(γ) *There exist open sets*‡ $N_1, N_2, N_3, \cdots$, *of which any open set is the sum of a subclass.*

HARVARD UNIVERSITY

† An open set is a set S such that, given $x\subset S$, g.l.b.$_{y\subset S}\overline{xy}>0$.

‡ An open set is a set S such that, if $\{x_s\}\subset\mathfrak{S}-S$, and $x\subset S$, then $\{x_s\}$ does not $\to x$

ON THE NUMBER OF STATIONARY TANGENT S_{t-1}'S TO A V_k^n IN AN S_{tk+k-1}

BY B. C. WONG

In this paper we propose to determine, geometrically, the number N of stationary tangent S_{t-1}'s to a k-dimensional variety V_k^n of order n in an S_{tk+k-1}. As the general problem of determining N for a V_k^n which is the locus of ∞^h S_{k-h}'s for $h>1$ offers some difficulty, we shall confine ourselves to the case where V_k^n is a non-developable locus of ∞^1 S_{k-1}'s. By a stationary tangent S_{t-1} to V_k^n we mean an S_{t-1} meeting V_k^n in $t+1$ consecutively coincident points, that is, meeting $t+1$ consecutive S_{k-1}'s of V_k^n each in a point.

Suppose that V_k^n belongs to an S_r. Now in S_r there are $\infty^{t(r+1-t)}$ S_{t-1}'s. For an S_{t-1} to meet V_k^n in $t+1$ points is equivalent to $(t+1, (r+1-k-t)$ conditions; and if the $t+1$ points of intersection are to be consecutively coincident, t further conditions will be absorbed. In order that the number N of stationary tangent S_{t-1}'s to V_k^n be finite, the dimension r of the space containing V_k^n must be such that $(t+1)(r+1-k-t)+t = t(r+1-t)$, from which we obtain $r=tk+k-1$. Our problem is to find N for V_k^n in S_{tk+k-1}.

Here we find it necessary to give two known results of which we shall make use subsequently.

(I) Let there be given q varieties $V_{x_1}^{m_1}$, $V_{x_2}^{m_2}$, $\cdots$, $V_{x_q}^{m_q}$ of orders $m_1, m_2, \cdots, m_q$, respectively, such that $V_{x_i}^{m_i}$ is the locus of ∞^1 (x_i-1)-spaces. If there exists a one-to-one correspondence between the elements of these varieties, then the locus of the ∞^1 $(x_1+x_2+\cdots+x_q-1)$-spaces determined by corresponding elements is a $V_{x_1+x_2+\cdots+x_q}$ of order $m_1+m_2+\cdots+m_q$.*

(II) The locus of the ∞^1 S_x's each meeting in $x+1$ coincident points a given curve of order m and deficiency p is a developable V_{x+1}^M of order $M=(x+1)(m-x+xp)$.†

* This result is a generalization of the proposition that the locus of the lines joining corresponding points of two projectively related curves is a ruled surface of order equal to the sum of the orders of the two curves.

† Veronese, *Behandlung der projectivischen Verhältnisse der Räume von verschiedenen Dimensionen durch das Princip des Projicirens und Schneidens*, Mathematische Annalen, vol. 19 (1882), pp. 161–234.

Now we proceed to the determination of N for V_k^n. We regard V_k^n as the projection of a $V_k'^n$ of the same order in an S_{tk+k+1} containing S_{tk+k-1}. Let there be given in S_{tk+k+1} k curves C^{m_1}, C^{m_2}, $\cdots$, C^{m_k} of orders $m_1, m_2, \cdots, m_k$, respectively, and all of the same deficiency p, and let a one-to-one correspondence exist between the points of these curves. Each group of corresponding points determines an S'_{k-1} and the locus of the ∞^1 S'_{k-1}'s so determined is, according to (I), a $V_k'^n$ of order $n=m_1+m_2+\cdots+m_k$. These S'_{k-1}'s will be called generating S'_{k-1}'s of $V_k'^n$ and the given curves will be called directrix curves, that is, curves each meeting each generating S'_{k-1} in a point.

Consider $t+1$ consecutive generating S'_{k-1}'s of $V_k'^n$. They determine a $(tk+k-1)$-space R_{tk+k-1}. The locus of the ∞^1 R_{tk+k-1}'s so determined is a V^ν_{tk+k} of order ν. To find ν, notice that an R_{tk+k-1} contains $t+1$ consecutive points on each of the k directrix curves and therefore the S_t determined by these $t+1$ points. The developable $V^{\mu_i}_{t+1}$ to the curve C^{m_i} is, according to (II), of order $\mu_i=(t+1)(m_i-t+tp)$. Then, from (I), the order of the locus V^ν_{tk+k} is $\nu=(t+1)(n-tk+tkp)$.

Now project $V_k'^n$ upon an S_{tk+k-1} from a general line l of S_{tk+k+1} and let the projection be V_k^n. The line l meets the locus V^ν_{tk+k} in ν points, that is, meets ν of the R_{tk+k-1}'s of V^ν_{tk+k}. Consider one of these R_{tk+k-1}'s and let the point of its incidence with l be P. From P one and only one S_t can be constructed meeting the $t+1$ consecutive S'_{k-1}'s of $V_k'^n$ lying in R_{tk+k-1} each in a point. Now an S_{t+1} containing l and S_t meets the S_{tk+k-1} upon which $V_k'^n$ is being projected in an S_{t-1}. This S_{t-1}, intersecting the projection V_k^n in $t+1$ consecutively coincident points which are the projections of the points on $t+1$ consecutive S'_{k-1}'s of $V_k'^n$, is a stationary tangent S_{t-1} of V_k^n. The number N of such stationary tangent S_{t-1}'s is evidently equal to the number of points in which l meets the locus V^ν_{tk+k} in S_{tk+k+1}, that is,

$$N = \nu = (t+1)(n-tk+tkp),$$

which was to be found.

We give a few illustrations of this formula. For $k=1$, we have a curve C^n in S_t, and the number of stationary tangent S_{t-1}'s to C^n is $N=(t+1)(n-t+tp)$. This particular result can be at once derived from (II) by projecting the curve upon a space of t dimensions. Now let $k=2$. If $t=1$, it follows that

$N=2(n-2+2p)$ is the number of pinch points on a ruled surface in S_3 and if $t=2$, then $N=3(n-4+4p)$ is the number of inflexional tangent lines to a ruled surface in S_5. We also see that a V_3^n, given by $k=3$, has $2(n-3+3p)$ pinch points if it is in S_5 and $3(n-6+6p)$ inflexional tangent lines if it is in S_8; and so forth.

THE UNIVERSITY OF CALIFORNIA

CLASS NUMBER IN A LINEAR ASSOCIATIVE ALGEBRA*

BY GRACE SHOVER

1. *Introduction.* In this paper the finiteness of the class number is established for every division algebra taken over the rational field. For every semi-simple algebra, the right and left class numbers are proved to be equal. The classical method for proving the finiteness of the class number in algebraic fields depends upon the multiplication of ideals, but the problem is treated in this paper for the general case without reference to the concept of ideal multiplication. The finiteness of the class number for every algebraic field follows as a special case.

2. *Definitions and References.* Algebra, domain of integrity, and ideal are defined as in a previous paper.†

The norm of an ideal $\mathfrak{R}$ is defined to be the absolute value of $|G|$, where G is a matrix representing $\mathfrak{R}$.‡

A necessary and sufficient condition that a matrix G represent a left (right) ideal is that

$$GR_p^T = D_pG, \ (GS_p = Q_pG), \quad (p = 1, 2, \cdots, n),$$

where R_p^T is the transpose of the first matrix of e_p (S_p is the second matrix of e_p). The matrices $D_p(Q_p)$ are called the class matrices corresponding to the ideal matrix G.

---

* Presented to the Society, November 28, 1931.

† Shover and MacDuffee, this Bulletin, vol. 37 (1931), pp. 434–438.

‡ MacDuffee, Transactions of this Society, vol. 31 (1929), pp. 71–90.

Every principal left (right) ideal may be represented by the second (transpose of the first) matrix of an integer of the albra. Every such matrix represents a left (right) ideal.

Two non-singular left (right) ideals $\mathfrak{K}_\alpha$ and $\mathfrak{K}_\beta$ are said to be equivalent if there exist integers of the domain σ_1 and σ_2 such that

$$\mathfrak{K}_\alpha\sigma_1 = \mathfrak{K}_\beta\sigma_2, \quad (\sigma_1\mathfrak{K}_\alpha = \sigma_2\mathfrak{K}_\beta).$$

In matric notation,

$$G_\alpha S(\sigma_1) = AG_\beta S(\sigma_2), \ (G_\alpha R^\tau(\sigma_1) = AG_\beta R^\tau(\sigma_2)),$$

where $|A| = \pm 1$. If $\mathfrak{K}_\alpha$ and $\mathfrak{K}_\beta$ are two equivalent non-singular left (right) ideals represented, respectively, by G_α and G_β, then the corresponding class matrices $D_{\alpha p}$ and $D_{\beta p}$ ($Q_{\alpha p}$ and $Q_{\beta p}$) are similar, that is,

$$AD_{\alpha p} = B_{\beta p}A, \ (AQ_{\alpha p} = Q_{\beta p}A),$$

where $|A| = \pm 1$, and conversely. If $D_{\alpha p} = D_{\beta p}$, $(Q_{\alpha p} = Q_{\beta p})$, G_α and G_β are said to belong to the same minor class. The left (right) class number is the number of non-equivalent non-singular left (right) ideals in the algebra.

It has been shown* that if T is the discriminant matrix of an algebra, the R and S matrices of a number are connected by the relation $ST = TR$.

3. *Class Number of Left Ideals.* Suppose

$$\beta = \sum b_i e_i, \ e_1 = 1,$$

is a number of the left ideal $\mathfrak{K}_\alpha$ with basis $\alpha_1, \alpha_2, \cdots, \alpha_n$. Then

$$\beta = \sum m_{1i}\alpha_i,$$

where the m_{1i} are rational integers. Multiplying by e_r on the left, we have

$$e_r\beta = \sum m_{1i}e_r\alpha_i, \qquad (r).$$

Since the α_i form a basis for a left ideal there exist rational integers m_{ri} such that

$$e_r\beta = \sum m_{ri}\alpha_i,$$

* MacDuffee, Transactions of this Society, vol. 33 (1931), pp. 425–432.

or

$$\sum b_i c_r e_i = \sum m_{ri} a_{is} e_s,$$

and

$$\sum b_i c_{ris} e_s = \sum m_{ri} a_{is} e_s.$$

Because of the linear independence of the basis numbers of the algebra, it follows that

$$\sum b_i c_{ris} = \sum m_{ri} a_{is}, \qquad (r, s).$$

Taking matrices, we have $S(\beta) = MG_\alpha$, where M is an integral matrix.

Conversely, if $S(\beta) = MG_\alpha$, then

$$e_r \beta = \sum m_{ri} \alpha_i, \qquad (r).$$

Let $r = 1$. Then

$$\beta = \sum m_{1i} \alpha_i.$$

We have now proved the following lemma for left ideals.

LEMMA 1. *If β is a number of the left ideal $\mathfrak{N}_\alpha$, there exists an integral matrix M such that $S(\beta) = MG_\alpha$, and conversely.*

LEMMA 2. *If β is any number not a divisor of zero of the non-singular left ideal $\mathfrak{N}$, then the adjoint M^A of $M = S(\beta)G^1$ represents a left ideal and it belongs to the same minor class as G.*

Since MG is a second matrix of the algebra, it is commutative with the transpose of every first matrix. Hence

$$MGR_p^T = R_p^T MG = MD_p G, \qquad (p).$$

Cancelling the G's in the last two members, multiplying them on both left and right by M^A, and dividing by $|M|$, we have

$$M^A R_p^T = D_p M^A.$$

LEMMA 3. *The determinants $|\Sigma h_i S_i|$, where the h_i are real and $|h_i| \leq 1$, have an upper bound in the algebra.*

The greatest number attainable for any determinant of order n is less than $n! g^n$, where g is the absolute value of an element of the determinant of maximum absolute value. But g can not be greater than $\Sigma |c_{ris}|$ for any r, s. Hence the determinant is bounded for any set of h_i's.

LEMMA 4. *The number of non-singular ideals whose norms are less than a given real number is finite.*

The number of ideals of any given norm g is finite. Applying Hermite's reduction theorem* to all matrices $G=(g_{rs})$ of norm equal to the absolute value g of $|G|$, we have

$$\pi g_{rr} = g, \qquad (g_{rr} > 0).$$

There are a finite number of choices for the g_{rr}. Since all the elements are positive and we have

$$g_{rs} < g_{rr}, \qquad (r > s),$$

and

$$g_{rs} = 0, \qquad (r < s),$$

there are a finite number of choices for all the elements in the matrix. Since there are a finite number of choices for norms less than a given real number, we have the following theorem.

THEOREM 1. *In a division algebra the left class number is finite.*

By a theorem of Minkowski,† it is possible to obtain integers x_i, $(i=1, 2, \cdots, n)$, not all zero such that

$$\left| \sum x_i a_{ij} \right| \leqq (|a|)^{1/n},$$

where $G=(a_{ij})$ is a matrix representing the non-singular left ideal $\mathfrak{A}_a=(\alpha_1, \alpha_2, \cdots, \alpha_n)$ and $a=|G|$. If $\beta=\Sigma x_i\alpha_i$, then

$$S(\beta) = MG_a \text{ and } S(\beta) = \sum x_i a_{ij}(c_{rjs}).$$

Since the algebra is a division algebra, $S(\beta)$ is non-singular and $|M| \neq 0$. Then

$$|S(\beta)| = \pm |G| \cdot \left| \sum h_i S_i \right|,$$

where

$$|h_i| = \left| \sum x_i a_{ij}/(|a|)^{1/n} \right| \leqq 1.$$

Since $|\Sigma h_i S_i|$ is bounded by a number b, $|M|$ is also bounded by b and $|M^A| \leqq b^{n-1}$. But M^A represents a non-singular ideal and it belongs to the same minor class as G. Hence in every class

* Journal für Mathematik, vol. 41 (1851), p. 193.

† Minkowski, *Geometrie der Zahlen*, Teubner, 1910, p. 104.

it is possible to find a non-singular ideal whose norm is less than a given finite number which depends only on the algebra. Therefore the class number is finite.

4. *Equality of the Two Class Numbers.*

LEMMA 5. *If G is a left ideal matrix, then $G^{AT}T^{A}$ is a right ideal matrix whose class matrices are the transposes of those of G.*

We have

$$GR_p^{T} = D_pG, \qquad (p),$$

where the D_p are integral. Also

$$S_pT = TR_p, \quad \text{or} \quad R_p^{T} = T^{T}S_p^{T}T^{T}.$$

Then

$$GT^{T}S_p^{T} = D_pGT^{T}, \quad \text{or} \quad S_pTG^{T} = TG^{T}D_p^{T}.$$

Hence

$$G^{AT}T^{A}S_p = D_p^{T}G^{AT}T^{A}.$$

THEOREM 2. *In every semi-simple algebra the left class number is equal to the right class number.*

Consider the left ideal matrices G_α and G_β and the corresponding right ideal matrices $G_\alpha^{AT}T^{A}$ and $G_\beta^{AT}T^{A}$. If the $D_{\alpha p}$ are (not) similar to the $D_{\beta p}$, the $D_{\alpha p}^{T}$ are (not) similar to the $D_{\beta p}^{T}$. Hence the right class number is not less (greater) than the left. Hence the two are equal.

Lemmas and theorems similar to those given here for left ideals may be proved for right ideals.

COLUMBUS, OHIO

CONCERNING COMPACT CONTINUA IN CERTAIN SPACES OF R. L. MOORE

BY J. H. ROBERTS

While writing his colloquium book, *Foundations of Point Set Theory*,† R. L. Moore noted that a large body of theorems concerning internal properties of compact continua could be established on the basis of a set of axioms (Axioms 1–5, *Foundations*) insufficient to make the space S itself a subset of a plane. He suggested that possibly every compact continuum M in S was homeomorphic with a compact continuum in the plane. In the present paper it is shown that, with possibly one exception, this is the case. If S is itself compact then it is homeomorphic with a subset of a sphere (possibly the sphere itself). But any compact continuum which is a *proper* subset of S is homeomorphic with a compact continuum in the plane.

THEOREM 1. *If S is a space in which Moore's Axioms* 1–5 *hold true and M is a closed and compact subset of S, then M is homeomorphic with a subset of a sphere. If furthermore M is a proper subset of S, then it is homeomorphic with a subset of a plane.*

Let E and J denote, respectively, a simple domain‡ and its boundary. It will be shown that if L is any circle in a plane and T_1 is any topological transformation of J into L, then there exists a topological transformation T_2 of $E \cdot M + J$ into a subset of L plus its interior, such that for each point P of J, $T_2(P) = T_1(P)$. From this result it readily follows that any closed and compact subset of S is homeomorphic with a subset of a sphere. If M is a closed and compact *proper* subset of S, and P is a point of $S - M$, then the closed and compact point set $M + P$ is homeomorphic with a subset of a sphere, and thus M is homeomorphic with a proper subset of a sphere, and hence homeomorphic with a subset of a plane. Henceforth it will be assumed that M is a closed and compact subset of $E + J$, where E is a simple domain and J is its boundary.

† Colloquium Publications of this Society, vol. XIII. Henceforth this book will be referred to as *Foundations*.

‡ A *simple domain* is a domain bounded by a simple closed curve.

The compact point set $M+J$ is completely separable (I, 19).† Then for every positive integer n there exists a countable subcollection of regions of G_n covering $M+J$ (I, 20). Thus there exists a countable collection G of regions such that for every n the set of regions of the collection G which are regions of G_n covers $M+J$. For each ordered pair (H, K) of regions of the set G for which there exists a pair (p, q) of simple domains such that‡ $H \supset [\bar{p} \cdot (M+J) +$the boundary of $p]$, $p \supset \bar{q}$, and $q \supset \overline{K}$, select some definite pair of simple domains with these properties. If R is any point of $M+J$, and H is any region of G containing R, then there exists a simple domain p containing R such that $H \supset [\bar{p} \cdot (M+J) +$the boundary of $p]$ (IV, 11). There exists a simple closed curve F lying in p and separating R from some point of $S-\bar{p}$. Let q denote the component of $S-F$ which contains R. Then $p \supset \bar{q}$ (III, 1). There exists a region K of the set G containing R and such that $q \supset \overline{K}$. Thus there exists a countable infinity of pairs (H, K) as described above, and with every such pair there can be associated a pair (p, q) of simple domains with properties as described above. Therefore there exists an infinite sequence of pairs (p_1, q_1), (p_2, q_2), $\cdots$, such that (1) for every n, p_n and q_n are simple domains and $p_n \supset \bar{q}_n$, and (2) if H and K are regions of the set G and there exist simple domains u and v such that $H \supset [\bar{u} \cdot (M+J) +$the boundary of $u]$, $u \supset \bar{v}$ and $v \supset \overline{K}$, then there exists an n such that $H \supset [\bar{p}_n \cdot (M+J) +$the boundary of $p_n]$ and $q_n \supset \overline{K}$.

Let A, B, C, and D denote distinct points of J in the order $ABCDA$. Let $P_1, P_2, \cdots$ denote an infinite sequence of points lying on $J-(A+B+C+D)$ and such that the set $P_1+P_2+\cdots$ is everywhere dense on J. Two lemmas will be interpolated at this point.

LEMMA 1. *If α is a double ruling§ of the interior of the simple closed curve $ABCDA$ and E_1 and E_2 are simple domains such that $E_1 \supset \overline{E}_2$, then there exists a simple closed curve L and a double*

† Unless otherwise stated, all references are to *Foundations*. (I, 19) denotes Theorem 19 of Chapter I.

‡ The statement "A contains B," or "B is a subset of A," is written $A \supset B$.

§ For a definition of this term see *Foundations*, pp. 404–405. The statement "α is a double ruling of the interior of $ABCDA$" is to be interpreted as implying that the arcs of one of the single rulings of α are parallel to AB and CD, and those of the other single ruling are parallel to AD and BC. Either complementary domain of $ABCDA$ may be regarded as its *interior*.

ruling β of the interior of $ABCDA$ such that (1) *L separates $\overline{E}_2$ and $S-E_1$*, (2) *every arc of α is also an arc of β, and* (3) *L is a subset of the sum of the arcs of β.*

Let F denote a finite point set containing A, B, C, and D, every point common to two arcs of α, and every end point of an arc of α. Let H denote $ABCDA$ plus all arcs of α. Let J_i denote the boundary of E_i $(i=1, 2)$. There exist mutually exclusive arcs R_1R_2 and S_1S_2 lying in $E_1-\overline{E}_2$ except that R_i and S_i are on J_i $(i=1, 2)$ (IV, 19 and II, 1). Then (III, 16) there exist domains U and V, arcs $R_1X_1S_1$ and $R_1Y_1S_1$ on J_1, and arcs $R_2X_2S_2$ and $R_2Y_2S_2$ on J_2 such that (1) U is bounded by $R_1X_1S_1+S_1S_2+R_2X_2S_2+R_1R_2$, V is bounded by $R_1Y_1S_1+S_1S_2+R_2Y_2S_2+R_1R_2$, and (2) $U+V+$segment R_1R_2+segment $S_1S_2=E_1-\overline{E}_2$. There exists a point $P(Q)$ on the segment R_1R_2 (S_1S_2) which does not belong to F and such that (1) either $P(Q)$ does not belong to H, or (2) $P(Q)$ belongs to an arc segment which is a subset of both H and R_1R_2 (S_1S_2). It can be shown that there exists an arc $PZ_1Q(PZ_2Q)$ which lies except for P and Q in $U(V)$, contains no point of F and at most a finite number of points of H, and such that if T is a point of H on the segment PZ_1Q (PZ_2Q), then PZ_1Q (PZ_2Q) *crosses at the point* T† that arc of α (or AB, BC, CD, DA) which contains T. Since Z_1P-P is in U and Z_2P-P is in $S-\overline{U}$, it follows that the arc Z_1PZ_2 crosses the arc R_1PR_2 at the point P. Then (IV, 32) R_1PR_2 crosses Z_1PZ_2 at P. Likewise S_1QS_2 crosses Z_1QZ_2 at Q. It follows that the simple closed curve PZ_1QZ_2P separates $\overline{E}_2$ and $S-E_1$. Furthermore, if P (Q) belongs to H then Z_1PZ_2 (Z_1QZ_2) crosses at P (Q) that arc of α (or AB, BC, CD, DA) which contains $P(Q)$.

Let L denote the simple closed curve PZ_1QZ_2P. Let J' denote the boundary of any of the subdivisions into which α divides the interior of $ABCDA$. Let A', B', C', and D' denote the points of F on J' so that on J' the order $A'B'C'D'A'$ obtains. Then the simple closed curve L satisfies, with respect to $A'B'C'D'A'$, the requirements‡ of Theorem 8 of Chapter VI of *Foundations*, and

† See *Foundations*, p. 201.

‡ Furthermore, if X is a point common to L and $A'B'C'D'A'$, then there exist arcs UXV and ZXY, being subsets of L and $A'B'C'D'A'$, respectively, which *cross* at the point X. Without this additional hypothesis, Theorem 8 of Chapter VI fails to hold true. Though stated in Chapter VI, the modified theorem holds true under Axioms 1–5.

there exists a double ruling $\gamma_{J'}$ of $A'B'C'D'A'$ such that every point of L within $A'B'C'D'A'$ lies on some arc of $\gamma_{J'}$. Thus there exists a double ruling of every subdivision into which α divides the interior of $ABCDA$. The arcs of these rulings can be extended so that there results a double ruling β of $ABCDA$ such that (1) every arc of α is also an arc of β, and (2) every arc of $\gamma_{J'}$ (for every J') is a subset of an arc of β. Then L is a subset of the sum of the arcs of β. This completes the proof of Lemma 1.

LEMMA 2. *If P is a point of a simple closed curve $ABCDA$ distinct from A, B, C, and D, and α is a double ruling of the interior of $ABCDA$, then there exists a double ruling β of the interior of $ABCDA$ such that* (1) *every arc of α is an arc of β, and* (2) *the point P is an end point of an arc of β.*

Returning to the proof of Theorem 1 we can show, using Lemmas 1 and 2 repeatedly, that there exists an infinite sequence $\alpha_1, \alpha_2, \cdots$ of double rulings of the interior of $ABCDA$ (consider E as the interior of $ABCDA$) such that, for every i, $(i=1, 2, \cdots)$, (1) every arc of α_i is also an arc of α_{i+1}, (2) there exists a simple closed curve L_i separating $\bar{q}_i$ and $S-p_i$ and being a subset of the sum of the arcs of α_i and (3) the point P_i is an end point of some arc of α_i.

There exists a homeomorphic correspondence between AD (AB) and the set of real numbers x $(0 \leqq x \leqq 1)$ such that A corresponds to 0 and D (B) corresponds to 1. If P is any point on AD (AB) let x_p (y_p) denote the number corresponding to P. If P is any point of $M+J$ and UV is an arc not containing P and belonging, for some i, to the double ruling α_i, then if UV is parallel to AB (AD) the point P is said to lie *to the right of* UV (*above* UV) if and only if it is true that of the mutually exclusive domains into which UV divides E, the one having on its boundary the point D (the point B) either contains P or has P on its boundary. Now let P denote any point of $M+J$. From the Dedekind-cut proposition (I, 64), it follows that there exists a unique point Q on AD such that P lies to the right of every arc of α_i $(i=1, 2, \cdots)$, which has an end point on the segment AQ, but does not lie to the right of any arc of α_i which has an end point on the segment QD. Define x_P to be the number x_Q. Similarly there exists a unique point R on AB such that P lies above every arc of α_i $(i=1, 2, \cdots)$, which has an end point on the segment AR, but does not lie above any arc of α_i

which has an end point on the segment RB. Define y_P to be the number y_R. Thus for every point P of $M+J$ there has been defined a pair of numbers (x_P, y_P). Let T denote the transformation of $M+J$ which throws the point P into the point with coordinates (x_P, y_P) in a plane. Then $T(J)$ is a square and $T(M)$ is a subset of a square plus its interior. It will now be shown that T is a topological transformation.

If it is shown that T is a one-to-one transformation, then it will follow that if the point P of $M+J$ is a limit point of the subset N of $M+J$, then $T(P)$ is a limit point of $T(N)$. To show that T is a topological transformation it will, then, be sufficient to show that if P is a point of $M+J$ and N is a closed subset of $M+J$ not containing P, then $T(P)$ is not a limit point of $T(N)$ and does not belong to $T(N)$. Let P denote a point of $M+J$ and N any closed subset of $M+J$ not containing P. Suppose W is a domain containing P. There exists a region H of the set G containing P, lying in W, and containing no point of N. There exists a simple domain u containing P such that $H \supset [\bar{u} \cdot (M+J) + \text{the boundary of } u]$. There exists a simple domain v containing P and such that $u \supset \bar{v}$. There exists a region K of the set G containing P and such that $v \supset \overline{K}$. Then for some integer (say i_1) the simple domains p_{i_1} and q_{i_1} have the properties stated above for u and v, respectively. Furthermore the simple closed curve L_i separates $\bar{q}_{i_1}$ and $S-P_{i_1}$ (and therefore K and $S-H$) and is a subset of the sum of the arcs of α_{i_1}. By a repetition of this argument it can be shown that there exists an integer i_2 $(i_2 > i_1)$, such that L_{i_2} separates P from L_{i_1}. Thus there exist two mutually exclusive simple closed curves, L_{i_1} and L_{i_2}, each separating P from N, and each being a subset of the sum of the arcs of α_{i_2}. There exists a positive number e such that if U_1V_1 and U_2V_2 are any arcs of α_{i_2} parallel to AB (AD) with U_1 and U_2 on AD (AB), then $|x_{U_1} - x_{U_2}| > e$ $(|y_{U_1} - y_{U_2}| > e)$. Clearly, then, if Q is any point of N, either $|x_P - x_Q| > e$ or $|y_P - y_Q| > e$. Then the distance from $T(P)$ to $T(Q)$ is greater than e. Hence T is a topological transformation of $M+J$ into a subset of a square plus its interior.

Let K denote the square which is the image of J, that is, $K = T(J)$. Let L denote any circle in the plane and suppose T_1 is a given topological transformation of J into L. Let T' denote the transformation of K into L such that, if P is a point of J,

so that $T(P)$ is a point of K and $T_1(P)$ is a point of L, then $T'[T(P)]=T_1(P)$. Clearly T' is a topological transformation. Then it can† be extended so that there results a topological transformation T'' of K plus its interior into L plus its interior such that on K, T'' reduces to T'. Now let P be any point of $M+J$. Let $T_2(P)$ denote the point $T''[T(P)]$. Then T_2 is a topological transformation of $M+J$ into a subset of the circle L plus its interior, and for each point P on J, $T_2(P)=T_1(P)$. This completes the proof of Theorem 1.

THEOREM 2. *If M is a compact continuum which is a proper subset of a space S in which Axioms* 1–5 *hold true, then there exists in a plane a compact continuum M^* homeomorphic with M.*

Theorem 2, which is an immediate consequence of Theorem 1, answers explicitly Moore's question mentioned in the introduction.

AXIOM 5′. *If P is a point of a domain D, then there exists a simple domain E containing P and lying in D.*

The set of Axioms 1, 3, 4, 5′ is clearly as strong as the set 1–5, as Axiom 2 is a consequence of Axioms 5′ and 4, and Axiom 5 is a consequence of Axiom 5′. The new set of axioms is, in fact, stronger, as there exist spaces satisfying Axioms 1–5 in which Axiom 5′ does not hold true. (See the Preface, *Foundations*.)

THEOREM 3. *If S is a space in which Axioms* 1, 3, 4, 5′ *hold true and M is a completely separable subset of S, then M is homeomorphic with a subset of a sphere. If furthermore M is a proper subset of S, then it is homeomorphic with a subset of a plane.*

The proof of Theorem 3 follows closely the proof of Theorem 1. In proving Theorem 1 the fact that M was closed and compact, rather than merely completely separable, was used only to show that if P is a point of M lying in a domain D then there exists a simple domain u containing P and such that $D \supset (\bar{u} \cdot M +$ the boundary of $u)$. But Axiom 5′ gives directly a stronger result than this.

† Schoenflies, *Beiträge zur Theorie der Punktmengen*, Mathematische Annalen, vol. 62 (1906), pp. 286–328. See also J. R. Kline, *A new proof of a theorem due to Schoenflies*, Proceedings of the National Academy of Sciences, vol. 6 (1920), pp. 529–531.

There exists a metric space S in which Axioms 1, 3, 4, and 5′ hold true, but such that S is not completely separable. Hence it does not follow that a space S in which Axioms 1, 3, 4, and 5′ hold true is a subset of a plane, even if it is assumed that S is metric.

DUKE UNIVERSITY

ANALOGS OF THE STEINER SURFACE AND THEIR DOUBLE CURVES*

BY A. R. WILLIAMS

The equations $x_1:x_2:x_3:x_4=x^n:y^n:z^n:w^n$, where x, y, z, w are linear functions of three homogeneous parameters, represent a rational surface of order n^2. For $n=2$ we have the well known Steiner surface. The particular subject of this paper is the double curve of such a surface and its representation on the plane. A few general properties must first be mentioned.

We take in the plane the reference system $x=0$, $y=0$, $z=0$, and $x+y+z\equiv -w=0$. The diagonals of the quadrilateral are $x+y\equiv -z-w=0$, etc. The vertices of the diagonal triangle are $(1:1:-1:-1)$, $(1:-1:1:-1)$, $(1:-1:-1:1)$, the fourth coordinate being w. Corresponding to the diagonals, the surface has 3 multiple right lines of order n, each meeting two opposite edges of the tetrahedron in points which correspond to a pair of opposite vertices of the quadrilateral. If n is even, the multiple lines are concurrent at $(1:1:1:1)$, which is a point of order $3(n-1)$ for the surface, corresponding to the vertices of the diagonal triangle and to certain pairs of imaginary points when $n>2$. If n is odd, the multiple lines are not concurrent, but are coplanar, meeting two by two at 3 points corresponding to the vertices of the diagonal triangle. The intersection of two multiple lines is then a point of order $2n-1$ for the surface. The class of the surface is always $3(n-1)^2$. The only pinch points are the 6 in which the multiple lines meet the edges of the tetrahedron. Each coordinate plane contains a single curve of order n, and is tangent to the surface along that curve, the order of contact being $n-1$. When n is even the section by a plane through a multiple line meets it in one variable real point, and

* Presented to the Society, March 18, 1933.

has at the node of the surface a point of order $2n-3$ with one real tangent. When n is odd the section by a plane through a multiple line has a point of order $n-1$ with imaginary tangents where the multiple line meets the other two. Of variable intersections with the multiple line two at most are real. In either case the plane determined by two multiple lines, or all three, contains in addition only isolated points of the surface.

The 3 double lines of the Steiner surface constitute its double curve. But for $n=3$ the order of the surface is 9, and we must have a double curve of order 27, since the genus of a plane section is 1. The 3 triple lines account for only 9. The plane representation indicates that the double curve, aside from the triple lines, consists of points isolated from the continuous portion of the surface and corresponding to pairs of conjugate imaginary points in the plane. To the section by the plane $x_1-x_2=0$ corresponds the plane curve $(y-x)(y-\omega x)(y-\omega^2 x)=0$. On the imaginary lines we may take $1:\omega:-t\omega^2:(1+t)\omega^2$ and $1:\omega^2:-t\omega:(1+t)\omega$, respectively, for $x:y:z:w$. These give $x_1:x_2:x_3:x_4=1:1:-t^3:(1+t)^3$, the parametric equations of a real plane cubic. For $t=\omega$ and $t=\omega^2$ the double cubic has the acnode $(1:1:-1:-1)$, which is the intersection of the two triple lines that do not meet $x_1=x_2=0$. There are evidently 6 such double cubics. The two that lie in planes through opposite edges of the tetrahedron have the same acnode.

Turning now to the general case, since $y-tx$ corresponds to a rational plane curve of order n in the plane $x_2-t^n x_1=0$, we infer that there may be $(n-1)(n-2)/2$ double curves of order n in planes through the opposite edge $x_3=x_4=0$. We shall find that this is so. But this is not enough. We have the relation

$$(n^2-1)(n^2-2)/2-(n-1)(n-2)/2=3n(n-1)/2+6n(n-1)(n-2)/2+n(n-1)(n-2)(n-3)/2.$$

The left side is the order of the double curve. The first term on the right is the equivalent of the three multiple lines; the second term is the portion just described; and the third term indicates that there are probably on the surface $(n-1)(n-2)(n-3)/2$ skew double curves each of order n. We shall find this to be the fact. It will be convenient to begin with n odd. Let $e^1, e^2, \cdots, e^{n-1}$ be the complex nth roots of unity, e being the root of smallest amplitude. Consider the equations

$$1 + re^{i} = se^{k},\ 1 + re^{n-i} = se^{n-k}, \tag{1}$$

where r and s are real. Either of these relations implies the other. The graphical interpretation is obvious. We readily obtain the relations $s/r=(e^{i}-e^{-i})/(e^{k}-e^{-k})$, $r=-(e^{k}-e^{-k})/(e^{k-i}-e^{-k+i})$, $s=-(e^{i}-e^{-i})/(e^{k-i}-e^{-k+i})$, where e^{-i} is written for e^{n-i}. There are $(n-1)(n-2)/2$ pairs of equations (1). Now we have in the plane the pair of conjugate points

$$\begin{aligned} x:y:z:w &= 1:t:(1+t)re^{i}:-(1+t)se^{k}, \\ x:y:z:w &= 1:t:(1+t)re^{-i}:-(1+t)se^{-k}, \end{aligned} \tag{2}$$

where t is a parameter. They give

$$x_1:x_2:x_3:x_4 = 1:t^{n}:(1+t)^{n}r^{n}:-(1+t)^{n}s^{n}. \tag{3}$$

If we take for t a fixed real number and use for i and k the $(n-1)(n-2)/2$ possible arrangements we get the acnodes of the rational curve of order n lying on the surface in the plane $x_2-t^{n}x_1=0$, whose image is $y-tx=0$. But if we take t variable, in general real, we get a double curve of order n on the surface in the plane $x_4/x_3+s^{n}/r^{n}=0$. Its parametric equations are (3). Its image in the plane is the line pair $se^{k}z+re^{i}w=0$ and $se^{-k}z+re^{-i}w=0$, or

$$s^{2}z^{2} + rs(e^{k-i} + e^{-k+i})zw + r^{2}w^{2} = 0. \tag{4}$$

If we set $t=-1$ in (3), we see that the double curve meets the edge $x_3=x_4=0$ where the latter is met by the multiple line. This double curve must itself be rational. Its nodes correspond to appropriate pairs of values of t in (3). Evidently these pairs are complex numbers and the nodes of the double curve are all acnodes. They are easily found from (1). For if we put for t in (3) the re^{i} and re^{-i} determined by the $(n-1)(n-2)/2$ pairs of equations (1), we shall have $(1+t)^{n}$ real. Furthermore it is easily shown that each acnode of the curve (3) is at the same time an acnode of a double curve lying in a plane through the opposite edge $x_1=x_2=0$. In other words, any pair of lines of the set (4) taken with any pair of similar character through $x=y=0$ give 4 imaginary points conjugate in pairs and corresponding to the same point of the surface which is an acnode for both double curves. The ratio $s/r=-1$ in the $(n-1)/2$ cases when $k+i=n$. Thus there are $(n-1)/2$ double curves in the *central* plane

$x_3-x_4=0$. Their images are $z^2-(e^{2i}+e^{-2i})zw+w^2=0$, $(i=1, 2, \cdots, (n-1)/2)$. Each plane double curve, therefore, has $(n-1)/2$ collinear acnodes on the intersection of its plane with the central plane through the opposite edge. If s/r is not -1, its negative as well as its reciprocal occurs. Moreover, the multiple line in $x_3+x_4=0$ is equivalent to the $(n-1)/2$ double curves in $x_3-x_4=0$. Thus the entire system of double curves in planes through an edge of the tetrahedron is symmetrically and harmonically placed with respect to the faces that intersect in that edge.

To find the skew double curves we use the equations

$$(5) \qquad 1+te^i+re^j=se^k,\ 1+te^{-i}+re^{-j}=se^{-k}$$

where t, r, and s are real. Solving for r and s we find $r=At+B$, $s=Ct+D$, where $A=-(e^{k-i}-e^{-k+i})/(e^{k-j}-e^{-k+j})$ and B, C, D are similar expressions having the same denominator as A. We then have as before

$$(6) \qquad \begin{aligned} x:y:z:w &= 1:te^i:(At+B)e^j:-(Ct+D)e^k, \\ x:y:z:w &= 1:te^{-i}:(At+B)e^{-j}:-(Ct+D)e^{-k}, \end{aligned}$$

which give

$$(7) \qquad x_1:x_2:x_3:x_4=1:t^n:(At+B)^n:-(Ct+D)^n,$$

the parametric equations of a skew curve of order n. To find its image we eliminate t by use of any two of the equations (6). Thus we find the line pair

(8) $Bx+Ae^{-i}y-e^{-j}z=0$ and $Bx+Ae^{i}y-e^{j}z=0$.

The product of the left hand members has of course real coefficients. The imaginary lines have one real point in common which is found by solving (8), and which may be obtained by putting $-e^{-i}BD/(AC)$ and $-e^{i}BD/(AC)$ for t in the first and second of equations (6). These values put for t in (7) give the same real point. Thus the skew double curve (7) has an acnode corresponding to the vertex of the imaginary line pair.

When n is even the result is the same, but there are certain modifications which deserve mention. If n is odd a number whose nth power is real must be the product of a real number and an nth root of unity. But if n is even the nth power of any $2n$th root of unity is real, that is, -1. Therefore, in forming equa-

tions (1) and (5) for n even we must use the $2n$th roots of unity. If e is the complex root of smallest amplitude, we have $e^1, e^2, \cdots, e^{2n-1}$, where $e^n = -1$. In the equations

$$(1') \qquad 1 + re^i = se^k,\ 1 + re^{2n-i} = se^{2n-k},$$

we may take i and k less than n since $e^{n+i} = -e^i$. Further, if we replace each root in (1') by its symmetric with respect to the imaginary axis, that is, replace i by $n-i$ and k by $n-k$, we merely change the signs of r and s and interchange the 2 imaginary lines which are the image of the plane double curve. Thus we have as before $(n-1)(n-2)/2$ pairs. Equation (4) for the image of a double curve remains the same, and $x_4/x_3 = (-se^k/(re^i))^n$ as before. It is easily seen that now s/r is always positive, and the minus sign has no effect since n is even; but $(e^k/e^i)^n$ is negative if i and k are of unlike parity; that is, if one is even and the other is odd. If e^i and e^k are symmetric with respect to the imaginary axis they are of the same parity, s/r becomes 1, and the corresponding plane curve lies in the *central* plane $x_3 - x_4 = 0$. There are $(n-2)/2$ such curves. Each has one of its acnodes at the intersection of the multiple lines. There is a further distinction. For n odd there are $(n-1)/2$ double curves in a *central* plane and all the other plane double curves are in separate planes. Now if in (1'), the exponent i being less than $n/2$ and $k \neq n/2$, we replace k by $n-k$, the ratio s/r and the parity will be unchanged, but the plane double curve obtained will be different. We have, therefore, $(n-2)(n-4)/2$ planes which contain 2 double curves. The point where such a plane meets a multiple line is an acnode on both curves, and a fixed quadruple point on the section by any plane through the multiple line. Such planes come in pairs. First interchanging i and k in (1'), and then replacing i by $n-i$, we get 2 curves of the same parity as the preceding, the ratio s/r being replaced by its reciprocal. That is, the ratio x_4/x_3 in the one plane is the reciprocal of the like ratio in the other. If we put $n/2$ for one of the exponents in (1') and then interchange them, we get a pair of planes related as above, and each containing one double curve. There are $(n-2)/2$ such pairs. Such a plane meets a multiple line in a point which is an acnode on the double curve and on the section by any plane through the multiple line. We may summarize

the above thus: If a plane $x_4 = Rx_3$ contains 1 or 2 double curves, $Rx_4 = x_3$ contains the same number. The reciprocal of any ratio x_4/x_3 occurs, but never the negative. The planes that contain the double curves are symmetrically but not harmonically placed with respect to the faces of the tetrahedron. To find the skew double curves we follow the method of equations (5). If we replace the three roots involved by their symmetrics with respect to the imaginary axis we get nothing new. Thus there are $(n-1)(n-2)(n-3)/2$ skew double curves.

In closing, the case $n=4$ may be mentioned since it is the first in which skew double curves occur. The planes $x_4+4x_3=0$, $4x_4+x_3=0$, and $x_4-x_3=0$, each contain a double quartic. Their images are $(z+w)^2+z^2=0$, $(z+w)^2+w^2=0$, and $z^2+w^2=0$. There are thus eighteen such curves, three through each edge. There are three skew double quartics whose images are the imaginary line pairs $(x-y)^2+(z-w)^2=0$, $(x-z)^2+(y-w)^2=0$, $(x-w)^2+(y-z)^2=0$. The real point of each pair is a vertex of the diagonal triangle and corresponds to the node of the surface. Therefore, each skew double curve has an acnode at the latter point. In fact one of them is given by $x_1:x_2:x_3:x_4=1:t^4:-(t-1)^4/4:-(t+1)^4/4$, and for $t=\pm i$ the right side becomes $(1:1:1:1)$. If one of these surfaces could be illuminated in a dark place we should see the double curves forming a thread-like pattern attached at certain points to the continuous portion.

THE UNIVERSITY OF CALIFORNIA

INTEGRAL FUNCTIONS OBTAINED BY COMPOUNDING POLYNOMIALS*

BY J. F. RITT

1. *Introduction.* We consider a sequence of polynomials $P_n(z)$, $(n=1, 2, \cdots)$, where the degrees of the P_n do not exceed a fixed integer m and where each P_n, ordered in ascending powers of z, starts with the term z. We shall study the sequence of polynomials $Q_n(z)$ defined by

$$(1) \qquad Q_1(z) = P_1(z);\ Q_{n+1}(z) = Q_n[P_{n+1}(z)], \qquad (n = 1, 2, \cdots),$$

and also the sequence of polynomials $R_n(z)$ defined by

$$(2) \qquad R_1(z) = P_1(z);\ R_{n+1}(z) = P_{n+1}[R_n(z)], \qquad (n = 1, 2, \cdots).$$

If the coefficients, after the first, in P_n, are sufficiently small, these sequences will converge to integral functions. For instance, $\sin z$ can be obtained, in many ways, as a limit of a sequence (1). In what follows, our chief object will be to establish conditions under which the sequences converge to integral functions.

2. *The Sequence of $Q_n(z)$.* Let

$$P_n(z) = z + a_{n2}z^2 + \cdots + a_{nm}z^m, \quad (n = 1, 2, \cdots),$$

where m is an integer independent of n.

THEOREM 1. *Let a convergent series of positive numbers,*

$$(3) \qquad c_1 + c_2 + \cdots + c_n + \cdots,$$

exist such that $|a_{ni}| < c_n$, for every n and for $i=2, \cdots, m$. Then the sequence of polynomials $Q_n(z)$ converges to an integral function, the convergence being uniform in every bounded domain.

PROOF. For every n,

$$(4) \qquad U_n(z) = z + c_n(z^2 + \cdots + z^m)$$

is a majorant of $P_n(z)$. Let

$$V_1 = U_1;\ V_{n+1} = V_n(U_{n+1}), \quad (n = 1, 2, \cdots).$$

* Presented to the Society, April 14, 1933.

Then V_n is a majorant of Q_n. Also, if we let

$$\alpha_n = c_n(z^2 + \cdots + z^m),$$

we have

$$\begin{aligned} V_{n+1} - V_n &= V_n(z + \alpha_{n+1}) - V_n \\ &= \frac{dV_n}{dz}\alpha_{n+1} + \frac{1}{2!}\frac{d^2V_n}{dz^2}\alpha_{n+1}^2 + \cdots, \end{aligned}$$

from which it follows easily that $V_{n+1} - V_n$ is a majorant of $Q_{n+1} - Q_n$. For every positive z, $V_{n+1}(z) > V_n(z)$. These considerations show that our theorem will be proved if we can show that the sequence of V_n converges for every positive z.

Let b be any positive number. Let

$$h = 2b + 4b^2 + \cdots + 2^{m-1}b^{m-1}. \tag{5}$$

Then the infinite product $(1+hc_1) \cdots (1+hc_n) \cdots$ converges. Let p be a fixed integer such that

$$(1 + hc_{p+1})(1 + hc_{p+2}) \cdots < 2. \tag{6}$$

Let

$$W_1 = U_{p+1};\ W_{n+1} = W_n(U_{p+n+1}), \quad (n = 1, 2, \cdots).$$

It will plainly suffice to show that the sequence of W_n converges for $z = b$. For any n, by (4) and (5),

$$U_{p+n}(b) < b(1 + hc_{p+n}),$$

so that, by (6), $U_{p+n}(b) < 2b$. Hence

$$\begin{aligned} U_{p+n-1}[U_{p+n}(b)] &= U_{p+n}(b)[1 + c_{p+n-1}(U_{p+n}(b) + \cdots)] \\ &< U_{p+n}(b)[1 + hc_{p+n-1}] \\ &< b(1 + hc_{p+n-1})(1 + hc_{p+n}), \end{aligned}$$

and the last quantity, by (6), is less than $2b$. Continuing in this fashion, we find that, for every n,

$$W_n < b(1 + hc_{p+1}) \cdots (1 + hc_{p+n}) < 2b.$$

This shows that the $W_n(b)$, which increase with n, approach a limit. The theorem is proved.

That the condition placed on the P_n is critical with respect to the convergence of the Q_n, is seen on taking $P_n = z + c_n z^m$ with

$c_n>0$ and (3) divergent. The coefficient of z^m in Q will be $c_1+\cdots+c_n$ and Q_n will tend towards infinity with n for every positive z.

The function $\sin z$ can be expressed as a limit of polynomials Q_n. Let

$$P_n(z) = z - \frac{4}{3^{2n+1}} z^3. \tag{7}$$

The formula

$$\sin z = 3 \sin \frac{z}{3} - 4 \sin^3 \frac{z}{3}$$

gives then

$$\sin z = Q_n(3^n \sin 3^{-n} z).$$

From (7) we see that the Q_n converge to an integral function. This integral function must be $\sin z$, since $3^n \sin 3^{-n} z$ approaches z as n increases.*

3. *The Sequence of* $R_n(z)$. We shall study the sequence of $R_n(z)$ defined by (2).

THEOREM 2. *Let the* $P_n(z)$ *all be of degree at most* $m>1$. *Let a sequence of positive numbers* c_n *exist such that*

$$\limsup_{n\to\infty} c_n^{1/m^n} < 1, \tag{8}$$

and such that, for every n, *the moduli of the coefficients of* $z^2, \cdots, z^m$ *in* P_n *are all less than* c_n. *Then the* $R_n(z)$ *converge to an integral function, the convergence being uniform in every bounded domain.*

PROOF. Let r be a number which lies between the two members of (8). Then, for n large,

$$z + r^{m^n}(z^2 + \cdots + z^m)$$

will be a majorant of $P_n(z)$. A fortiori, since $m>1$,

$$U_n(z) = z + r^{m^n-1} z^2 + r^{2m^n-1} z^3 + \cdots + r^{(m-1)m^n-1} z^m \tag{9}$$

* In the same way, one can express as limits of polynomials Q_n a large class of the functions with rational multiplication theorems introduced by Poincaré (Journal de Mathématiques, vol. 55 (1890)).

will be a majorant of $P_n(z)$ for n large. We see now readily that it will suffice, for the proof of our theorem, to show that the sequence of $V_n(z)$ defined by

$$V_1 = U_1;\ V_{n+1} = U_{n+1}(V_n), \qquad (n \geqq 1), \tag{10}$$

converges for every real and positive z.†

Let p be any non-negative integer. Putting

$$W_1 = U_{p+1};\ W_{n+1} = U_{p+n+1}(W_n), \qquad (n \geqq 1), \tag{11}$$

we shall show that the sequence of W_n converges for $z < hr^{-m^p}$, where $h = 1 - r$.

By (9),

$$S_n(z) = \frac{z}{1 - r^{m^{p+n-1}}z}, \qquad (n = 1, 2, \cdots),$$

is a majorant of U_{p+n}. If, then,

$$T_1 = S_1;\ T_{n+1} = S_{n+1}(T_n), \qquad (n \geqq 1),$$

T_n will be a majorant of W_n. Now an easy calculation shows that

$$T_n(z) = \frac{z}{1 - (r^{m^p} + \cdots + r^{m^{p+n-1}})z}.$$

For any positive z less than the reciprocal of the infinite series

$$r^{m^p} + r^{m^{p+1}} + \cdots,$$

which reciprocal we shall denote by k, the $T_n(z)$ form a sequence of numbers which increase towards $kz/(k-z)$. Also, if $0 < z < k$, $T_n(z) > W_n(z)$, so that the $W_n(z)$ will form a bounded sequence of increasing numbers and will converge to a limit. Now as $m > 1$,

$$k \geqq \frac{r^{-m^p}}{1 + r + r^2 + \cdots} = hr^{-m^p},$$

and our statement with respect to (11) is proved.

Thus Theorem 2 will be established if, putting $V_0(z) = z$, we show that for every positive z there is a p such that $V_p(z) < hr^{-m^p}$.

† The fact that U_n may not be a majorant of P_n for n small is of no significance. One may suppress a finite number of P_n and then add a finite number of polynomials (9) to the beginning of the resulting sequence of U_n.

Let us assume that there is a positive z for which no such p exists. In what follows, we work with a fixed z of this type. We have, by (9) and (10), for any $n \geqq 0$,

$$V_{n+1} = V_n + r^{m^n} V_n^2 + \cdots + r^{(m-1)m^n} V_n^m \leqq V_n(1 + r^{m^n} V_n)^{m-1}. \tag{12}$$

Now, for every n,

$$V_n \geqq h r^{-m^n}, \tag{13}$$

so that

$$1 \leqq \frac{r^{m^n} V_n}{h},$$

and, if we put $a = (1+1/h)^{m-1}$, we have, by (12),

$$V_{n+1} \leqq a r^{(m-1)m^n} V_n^m \leqq a r^{m^n} V_n^m.$$

We have thus

$$V_1 \leqq arz^m, \quad V_2 \leqq a^{m+1} r^{2m} z^{m^2}, \quad V_3 \leqq a^{m^2+m+1} r^{3m^2} z^{m^3},$$

and, in general,

$$V_{n+1} \leqq a^{m^n+\cdots+1} r^{(n+1)m^n} z^{m^{n+1}}.$$

As $m > 1$, we have $m^n + \cdots + 1 < m^{n+1}$. Then, because $a > 1$,

$$V_{n+1} < \left[r^{(n+1)/m} az\right]^{m^{n+1}}. \tag{14}$$

As z is fixed, $r^{(n+1)/m} az$ is small for n large, so that, by (14), V_{n+1} approaches 0 as n increases. This contradicts (13). The theorem is proved.

The condition (8) is a critical one. That we cannot let the first member of (8) be as great as unity is seen on taking $P_n = z + z^m$. The coefficient of z^m in Q_n will be n and the Q_n will diverge for every positive z. That m in the first member of (8) cannot be replaced by any smaller positive number α, is seen, taking $m = 2$, for instance, on putting $P_n = z + 2^{-\alpha^n} z^2$. For any positive z, we have

$$P_n > 2^{-\alpha^n} z^2.$$

Then

$$R_1 > 2^{-\alpha} z^2, \quad R_2 > 2^{-(\alpha^2+2\alpha)} z^4,$$

and, in general,

$$R_n > 2^{-(\alpha^n+2\alpha^{n-1}+\cdots+2^{n-1}\alpha)}z^{2^n}.$$

Now

$$-(\alpha^n + 2\alpha^{n-1} + \cdots + 2^{n-1}\alpha) = \frac{\alpha^{n+1} - 2^n\alpha}{2-\alpha} > -b2^n,$$

where $b=\alpha/(2-\alpha)$. Thus

$$R_n > \left(\frac{z}{2^b}\right)^{2^n},$$

so that the R_n diverge for $z>2^b$.

Let $f(z)$ be an integral function obtained as a limit of polynomials $R_n(z)$, the approach being uniform in every bounded domain. Unless $P_n(z)=z$ for every n, $f(z)$ will not be linear, for if some $R_n(z)$ is of degree greater than unity, $f(z)$, like that $R_n(z)$, will assume certain values at more than one place. In what follows, we shall assume that $f(z)$ is not linear.

We are going to prove that, between any two branches of the inverse of $f(z)$, there exists an algebraic relation of a simple type.

Let a and b be two distinct points such that $f(a)=f(b)$ and that the derivative of $f(z)$ does not vanish at a or at b. Let A be a circle with a as center such that, in the interior of A, $f(z)$ assumes no value twice. Let B be a similar circle with center at B. We can find a neighborhood M of $f(a)=f(b)$ such that, both in A and in B, $R_n(z)$ with n large assumes all values in M. If n is large enough, $R_n(a)$ will be in M. In what follows, we deal with a fixed $R_n(z)$ for which both conditions just described are realized.

If z_a is a point in A, very close to a, there will be a z_b in B such that $f(z_b)=f(z_a)$, and, furthermore, $R_n(z_a)$ will lie in M. We shall prove that $R_n(z_a)=R_n(z_b)$. As $R_n(z_a)$ is in M, there is a ζ in B such that $R_n(\zeta)=R_n(z_a)$. Now ζ must coincide with z_b, for $f(\zeta)=f(z_a)=f(z_b)$ and $f(z)$ assumes no value twice in B.

Thus, *if we put* $w=f(z)$ *and if* $\alpha(w)$ *and* $\beta(w)$ *are two branches of the inverse of* $f(z)$, *then, for* n *large,* $R_n[\alpha(w)]=R_n[\beta(w)]$.

COLUMBIA UNIVERSITY

OFFICIAL COMMUNICATIONS

Meetings of the Society have been fixed at the following times and places:

NEW YORK CITY, October 28, 1933.

Abstracts must be in the hands of Associate Secretary J. R. Kline, 501 West 116th St., New York City, not later than October 7. Professor R. L. Jeffery will deliver an address on *Theories of integration*. Abstracts received by the Secretaries by September 9 will appear in the September issue of the BULLETIN.

CINCINNATI, OHIO, December 1-2, 1933.

Abstracts must be in the hands of Associate Secretary M. H. Ingraham, University of Wisconsin, Madison, Wis., not later than November 11. Abstracts received by the Secretaries by November 4 will appear in the November issue of the BULLETIN.

PASADENA, CALIFORNIA, December 2, 1933.

Abstracts must be in the hands of Associate Secretary T. M. Putnam, University of California, Berkeley, Calif., not later than November 11. Abstracts received by the Secretaries by November 4 will appear in the November issue of the BULLETIN.

CAMBRIDGE, MASSACHUSETTS, ANNUAL MEETING, December 27-29, 1933.

This meeting is to be held in conjunction with the A.A.A.S. and the M.A.A. There will be a symposium on *General Analysis*. The Bôcher Prize will be awarded. Abstracts must be in the hands of the Secretary of the Society, 501 West 116th St., New York City, not later than November 30, 1933. Abstracts received by the Secretaries by November 4 will be published in the November issue of the BULLETIN.

R. G. D. RICHARDSON, *Secretary of the Society.*

Articles for insertion in the BULLETIN should be addressed to E. R. HEDRICK, Editor of the BULLETIN, University of California at Los Angeles. Reviews should be sent to W. R. LONGLEY, Yale University, New Haven, Conn. Notes should be sent to H. W. KUHN, Ohio State University, Columbus, Ohio.

Subscriptions to the BULLETIN, orders for back numbers, and inquiries in regard to non-delivery of current numbers should be addressed to the American Mathematical Society, 450-459 Ahnaip St., Menasha, Wis., or 501 West 116th St., New York.

The initiation fees and the annual dues of members of the Society (see this BULLETIN, p. 322, May, 1930; and the List of Officers and Members, September, 1932, p. 66), are payable to the Treasurer of the Society, Professor G. W. Mullins, 501 West 116th St., New York City.

WHOLE NUMBER 407

CONTENTS

PAGE

On the Infinite Sequences arising in the Theories of Harmonic Analysis, of Interpolation, and of Mechanical Quadratures. By LEOPOLD FEJÉR 521

Some Mathematical Aspects of the New Mechanics. By TULLIO LEVI-CIVITA 535

Matrices with Elements in a Principal Ideal Ring. By C. C. MACDUFFEE 564

A Note on the Dickson Theorem on Universal Ternaries. By A. A. ALBERT 585

A Self-Dual Septimic possessing Seven of each Kind of the Simple Singularities, and Autopolar by seven Rectangular Hyperbolas and a Circle. By D. C. DUNCAN 589

On Irreducible Systems of Algebraic Differential Equations. By E. GOURIN 593

Axiom C of Hausdorff and the Property of Borel-Lebesgue. By SELBY ROBINSON 595

Axiomatic Definitions of Perfectly Separable Metric Spaces. By GARRETT BIRKHOFF 601

On the Number of Stationary Tangent S_{t-1}'s to a V_t^s in an S_{k+s-1}. By B. C. WONG 608

Class Number in a Linear Associative Algebra. By GRACE SHOVER 610

Concerning Compact Continua in certain Spaces of R. L. Moore. By J. H. ROBERTS 615

Analogs of the Steiner Surface and their Double Curves. By A. R. WILLIAMS 621

Integral Functions obtained by compounding Polynomials. By J. F. RITT 627

For official Communications and Notices, see the inside of the back cover.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN